

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 11:24:03 ; Search time 95.25 Seconds

(without alignments)  
294.716 Million cell updates/sec

Title: PCT-US02-13994-20  
Sequence: 3685  
1 MAMDSIQAARLPGALIRKIO.....LAQLEQASQISKKRPQ 725

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3683	99.9	725	1	MCAR_HUMAN
2	3252	88.2	718	1	MCAR_CRIGR
3	2856	77.5	671	1	KRP2_RAT
4	2361	64.1	730	1	KCM1_XENLA
5	1868.5	50.7	679	1	KIFP_HUMAN
6	1859	50.4	716	1	KIFP_MOUSE
7	1822	49.4	682	1	KIF2_XENLA
8	750	20.4	624	1	DSK1_CYLFU
9	574.5	15.6	1232	1	KFAA_HUMAN
10	571	15.4	1231	1	KFAA_MOUSE
11	569	15.4	1225	1	KFAA_CHICK
12	561.5	15.2	1056	1	K125_ARATH
13	552.5	15.0	1006	1	K125_TOBAC
14	548.5	14.9	1226	1	KFAA_XENLA
15	519	14.1	883	1	KLP3_SCHPO
16	515	14.0	702	1	KFAA_HUMAN
17	512.5	13.9	1067	1	EG53_XENLA
18	509.5	13.8	701	1	KFAA_MOUSE
19	506	13.7	1060	1	EG53_XENLA
20	504	13.6	1057	1	EG5_HUMAN
21	502	13.6	1111	1	KIF1_YEAST
22	498	13.5	805	1	KIF6_MOUSE
23	493.5	13.4	956	1	KIF6_MOUSE
24	491	13.3	957	1	KIF5C_HUMAN
25	491	13.3	1690	1	KIF5C_HUMAN
26	490	13.3	1066	1	KL61_DROME
27	489.5	13.3	1103	1	KL61_DROME
28	488	13.2	689	1	K122_STRPU
29	488	13.2	1695	1	KIFP_MOUSE
30	478.5	13.0	955	1	KIFP_MOUSE
31	478	13.0	786	1	KIFL_LEICH
32	477	12.9	1816	1	KIFB_MOUSE
33	473	12.8	747	1	KIFB_MOUSE

34	472	12.8	747	1	KIFB_HUMAN
35	468	12.6	742	1	K121_STRPU
36	465.5	12.6	1584	1	D104_CAEEL
37	464.5	12.6	1097	1	KIFD_RAT
38	464	12.6	963	1	KIFN_MOUSE
39	463.5	12.6	784	1	KL68_DROME
40	463	12.6	796	1	KIFC_MOUSE
41	462.5	12.6	1184	1	BIMC_EMENT
42	458	12.4	796	1	KIFC_MOUSE
43	458	12.4	1816	1	KIFB_MOUSE
44	457.5	12.4	967	1	KIFN_LOLPE
45	457	12.4	975	1	KIFN_DROME

## ALIGNMENTS

RESULT 1  
ID MCAR\_HUMAN STANDARD: PRT; 725 AA.  
AC 09661; 096C18; 096HB8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Mitotic centromere-associated kinesin (MCAR) (kinesin-like protein 6).  
GN KNSL6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-T-cell;  
RX MEDLINE:98094213; PubMed-9434124;  
RA Kim I.-G., Jun D.Y., Sohn U., Kim Y.H.;  
RT Cloning and expression of human mitotic centromere-associated kinesin  
RT gene.  
RL Blochim. Biophys. Acta 1359:181-186(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Kidney, and uterus;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
CC - FUNCTION: PRESENT THROUGHOUT THE CELL CYCLE. ASSOCIATES WITH  
CC CENTROMERES AT EARLY PROPHASE, AND REMAINS ASSOCIATED WITH THE  
CC CENTROMERE UNTIL AFTER TELOPHASE (BY SIMILARITY).  
CC - TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THYMUS AND TESTIS,  
CC AT LOW LEVELS IN SMALL INTESTINE, THE MUCOSAL LINING OF COLON, AND  
CC PLACENTA. AND AT VERY LOW LEVELS IN SPLEEN AND OVARY. EXPRESSION  
CC IS NOT DETECTED IN PROSTATE, PERIPHERAL BLOOD LEUKOCYTES, HEART,  
CC BRAIN, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, OR PANCREAS.  
CC - SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAR/KIF2  
CC SUBFAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; U63743; AAC27660.1;  
CC EMBL; BC014924; AAH14924.1;  
CC EMBL; BC008764; AAH08764.1;  
CC HSSP; P17119; 3KAR.  
CC MIM; 604538;  
CC InterPro: IPRO01752; kinesin.  
CC Pfam: PF00225; kinesin\_2.  
CC PRINTS: PR00380; KINESINHEAVY.  
CC SMART: SM00129; Kisc; 1.  
CC PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.

DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 KW Motor protein; Microtubules; ATP-binding; Coiled coil;  
 FT DOMAIN 1 252 GLOBULAR (POTENTIAL).  
 FT DOMAIN 253 618 KINESIN-MOTOR (BY SIMILARITY).  
 FT DOMAIN 619 658 COILED COIL (POTENTIAL).  
 FT DOMAIN 659 725 COILED COIL (POTENTIAL).  
 FT NP\_BIND 348 355 ATP (POTENTIAL).  
 FT DOMAIN 415 418 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 449 449 I -> L (IN REF. 1).  
 FT CONFLICT 698 698 R -> P (IN REF. 2; AA08764).  
 SO SEQUENCE 725 AA: 81312 MW: 5BDEC133AB4B55C CRC64.

Query Match 99.98; Score 3683; DB 1; Length 725;  
 Best Local Similarity 99.98; Pred. No. 1.8e-209;  
 Matches 724; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MANDSSIAQRLPFGIATKIOKNSGLHSANVTYNLEKSCSVENMAEGATGKEIDPD 60  
 1 MANDSSIAQRLPFGIATKIOKNSGLHSANVTYNLEKSCSVENMAEGATGKEIDPD 60  
 61 VAINPDLQLLPKPNLPLEQENTYIOKRRSVNSKIPAKRESLRSTRMSTVSEL 120  
 61 VAINPDLQLLPKPNLPLEQENTYIOKRRSVNSKIPAKRESLRSTRMSTVSEL 120  
 121 RITAQNDMEVELPAANSKROFVPAAPTRPSCPAVAELPLMNSSEMEEOVHSIRGS 180  
 121 RITAQNDMEVELPAANSKROFVPAAPTRPSCPAVAELPLMNSSEMEEOVHSIRGS 180  
 121 RITAQNDMEVELPAANSKROFVPAAPTRPSCPAVAELPLMNSSEMEEOVHSIRGS 180  
 181 SANPVSVRKSCQVKEVEKMKNEKKAKONSEMRKRAOEDSSFPNMEFARMKEFR 240  
 181 SANPVSVRKSCQVKEVEKMKNEKKAKONSEMRKRAOEDSSFPNMEFARMKEFR 240  
 241 ATECHPLMTPTPIEHRCVCVRRKPLNKELEKIDVISTPSCILLVHEPKAKYL 300  
 241 ATECHPLMTPTPIEHRCVCVRRKPLNKELEKIDVISTPSCILLVHEPKAKYL 300  
 241 ATECHPLMTPTPIEHRCVCVRRKPLNKELEKIDVISTPSCILLVHEPKAKYL 300  
 301 TKLENAQCFPFADETASNEVYRFTAPLVOTIFEGCATCFAYGOTSGKRTMCG 360  
 301 TKLENAQCFPFADETASNEVYRFTAPLVOTIFEGCATCFAYGOTSGKRTMCG 360  
 301 TKLENAQCFPFADETASNEVYRFTAPLVOTIFEGCATCFAYGOTSGKRTMCG 360  
 361 DLSGKONASKGYANASADVFLNOCYKRLLEVYVFEFYNNKLDLNTAKLR 420  
 361 DLSGKONASKGYANASADVFLNOCYKRLLEVYVFEFYNNKLDLNTAKLR 420  
 361 DLSGKONASKGYANASADVFLNOCYKRLLEVYVFEFYNNKLDLNTAKLR 420  
 421 VLEDGKQOVVVGLOEHLVNSADVIKMLDNGSACRTSGOTFANSNSRSRSHACFOITLRA 480  
 421 VLEDGKQOVVVGLOEHLVNSADVIKMLDNGSACRTSGOTFANSNSRSRSHACFOITLRA 480  
 421 VLEDGKQOVVVGLOEHLVNSADVIKMLDNGSACRTSGOTFANSNSRSRSHACFOITLRA 480  
 481 KGRHNGCFSLVDLAGNERADOTSSADROTMRGAEINKSLALKECIRALGOKRATPPR 540  
 481 KGRHNGCFSLVDLAGNERADOTSSADROTMRGAEINKSLALKECIRALGOKRATPPR 540  
 481 KGRHNGCFSLVDLAGNERADOTSSADROTMRGAEINKSLALKECIRALGOKRATPPR 540  
 541 ESKLTQVLRDSFGENSRTCMATISPGISCEYTLNTRADRVKELSPHSPEEQIL 600  
 541 ESKLTQVLRDSFGENSRTCMATISPGISCEYTLNTRADRVKELSPHSPEEQIL 600  
 541 ESKLTQVLRDSFGENSRTCMATISPGISCEYTLNTRADRVKELSPHSPEEQIL 600  
 601 OMETEMEACSNGLPGULSKBEELSSQSSFNEMATOIRELEFAMKELKEIIIOGP 660  
 601 OMETEMEACSNGLPGULSKBEELSSQSSFNEMATOIRELEFAMKELKEIIIOGP 660  
 601 OMETEMEACSNGLPGULSKBEELSSQSSFNEMATOIRELEFAMKELKEIIIOGP 660  
 661 DMLELSMTROPDYDLETPVNAKESALAOQAHFALADVAKALRLMOLEBOASROISS 720  
 661 DMLELSMTROPDYDLETPVNAKESALAOQAHFALADVAKALRLMOLEBOASROISS 720  
 661 DMLELSMTROPDYDLETPVNAKESALAOQAHFALADVAKALRLMOLEBOASROISS 720  
 721 KRPQ 725  
 721 KRPQ 725  
 721 KRPQ 725

RESULT 2  
 MCAK\_CRIGR

ID MCAK\_CRIGR STANDARD; PRT; 718 AA.  
 AC P70096;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Mitotic centromere-associated kinesin (MCAK) (kinesin-like protein 6).  
 GN KNSL6.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetus.  
 OX NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95122643; Pubmed=7822426;  
 RA Wordeman L., Mitchison T.J.;  
 RT Identification and partial characterization of mitotic centromere-  
 RT associated kinesin, a kinesin-related protein that associates with  
 RT centromeres during mitosis.  
 RL J. Cell Biol. 128:95-105(1995).  
 RN [2]  
 RP REVISIONS.  
 RA Wordeman L.;  
 RT Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: PRESENT THROUGHOUT THE CELL CYCLE. ASSOCIATES WITH  
 CC CENTROMERES AT EARLY PROPHASE, AND REMAINS ASSOCIATED WITH THE  
 CC CENTROMERE UNTIL AFTER TELOPHASE.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/11F2  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: U11790; AB17358.2;  
 CC HSP: P17119; 3KAR.  
 CC InterPro: IPR001752; Kinesin.  
 CC Pfam: PF00225; Kinesin.  
 CC PRINTS: PR00380; KINESINHEAVY.  
 CC SMART: SM00129; KISC.1.  
 CC DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 CC DR PROSITE: PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 KW Motor protein; Microtubules; ATP-binding; Coiled coil;  
 KM Nuclear protein.  
 FT DOMAIN 1 246 GLOBULAR (POTENTIAL).  
 FT DOMAIN 247 612 KINESIN-MOTOR (BY SIMILARITY).  
 FT DOMAIN 613 651 COILED COIL (POTENTIAL).  
 FT DOMAIN 659 716 COILED COIL (POTENTIAL).  
 FT NP\_BIND 342 349 ATP (POTENTIAL).  
 FT DOMAIN 409 412 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 SO SEQUENCE 718 AA: 80918 MW: 16ABD8C665AD1B2 CRC64.

Query Match 88.28; Score 3252; DB 1; Length 718;  
 Best Local Similarity 88.58; Pred. No. 4e-184;  
 Matches 637; Conservative 34; Mismatches 45; Indels 4; Gaps 2;

6 SIQARLPFGIATKIOKNSGLHSANVTYNLEKSCSVENMAEGATGKEIDPDVAAIN 65  
 6 SIQARLPFGIATKIOKNSGLHSANVTYNLEKSCSVENMAEGATGKEIDPDVAAIN 65  
 3 SIQARLPFGIATKIOKNSGLHSANVTYNLEKSCSVENMAEGATGKEIDPDVAAIN 62  
 66 PELLQLLPKPNLPLEQENTYIOKRRSVNSKIPAKRESLRSTRMSTVSELRTAQ 125  
 66 PELLQLLPKPNLPLEQENTYIOKRRSVNSKIPAKRESLRSTRMSTVSELRTAQ 125  
 66 PELLQLLPKPNLPLEQENTYIOKRRSVNSKIPAKRESLRSTRMSTVSELRTAQ 125  
 63 PELLQLLPKPNLPLEQENTYIOKRRSVNSKIPAKRESLRSTRMSTVSELRTAQ 122  
 126 ENDEVELEPAANSKROFVPAAPTRPSCPAVAELPLMNSSEMEEOVHSIRGS 185  
 126 ENDEVELEPAANSKROFVPAAPTRPSCPAVAELPLMNSSEMEEOVHSIRGS 185  
 123 ENDEVELEPAANSKROFVPAAPTRPSCPAVAELPLMNSSEMEEOVHSIRGS 181

QY	121	PIGCVSAAVENKPNVAMTISOPICAKAFITDIEDIRROERVOYRKKULEEALMADILSR	100
DB	124	PIGCVSAAVENKPNVAMTISOPICAKAFITDIEDIRROERVOYRKKULEEALMADILSR	100
QY	101	AADTEEMDIEDMSGDEA	197
DB	104	AADTEEMDIEDMSGDEA	200
RESULT	2		
Q9UBB5			
ID	Q9UBB5	PRELIMINARY;	PRT; 411 AA.
OC	Q9UBB5		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	MEHTYL-CPG BINDING PROTEIN 2.		
GN	MBD2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=99373255; PubMed=10441743;		
RA	Hendrich B., Abdoott C., McQueen H., Chambers D., Cross S., Bird A.;		
RT	"Genomic structure and chromosomal mapping of the mouse and human		
RL	mbd1, mbd2, mbd3, and mbd4 genes.";		
RL	Mamm. Genome 10:906-912(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=98449942; PubMed=9774669;		
RA	Hendrich B., Bird A.;		
RT	"Identification and characterization of a family of mammalian methyl-		
RT	Cpg binding proteins and		
RL	Mol. Cell. Biol. 18:6538-6547(1998).		
DR	EMBL; AF120993; AAD56597.1;		
DR	EMBL; AF120988; AAD56597.1;		
DR	EMBL; AF120989; AAD56597.1;		
DR	EMBL; AF120990; AAD56597.1;		
DR	EMBL; AF120992; AAD56597.1;		
DR	EMBL; AF072242; AAC68871.1;		
DR	InterPro: IPR001739; MBD.		
DR	Pfam: PF01429; MBD. 1.		
DR	SMART: SM00391; MBD. 1.		
QY	SEQUENCE 411 AA: 43224 MW: FC4E5E0CF9BA0F6A CRC64;		

Query Match	100.0%	Score 998	DB 4	Length 411	
Best Local Similarity	100.0%	Pred. No. 2,6e-76			
Matches 197	Conservative	0	Mismatches	0	Indels
				0	Gaps
					0
QY	1	PSKLOKKKRLNDPLNONGKGRDPLNTLPIRQTSIRKQVTVYTNHPSKRVASDQRM	60		
Db	215	PSKLOKKKRLNDPLNONGKGRDPLNTLPIRQTSIRKQVTVYTNHPSKRVASDQRM	274		
QY	61	NEOPROLFWKRLQGLSASDVTBOIITKTNELPKGLQGVGPGSNDETLLSAVASALHTSSA	120		
Db	275	NEOPROLFWKRLQGLSASDVTBOIITKTNELPKGLQGVGPGSNDETLLSAVASALHTSSA	334		
QY	121	PITGOVSAAEKKNPAWMLNTSOPLCRAFIYTDIEDIRKQEEFVQVRKKLEALMAADILSR	180		
Db	335	PITGOVSAAEKKNPAWMLNTSOPLCRAFIYTDIEDIRKQEEFVQVRKKLEALMAADILSR	394		
QY	181	AAOTFEEDIDEDSGDEA	197		
Db	395	AAOTFEEDIDEDSGDEA	411		
RESULT	3				
Q972E1		PRELIMINARY:	PRT	414 AA.	
Q972E1					

AC 09322E1;  
 DT 01-MAY-1999 (TRENBLREL. 10, Created)  
 DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)  
 DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)  
 DE METHYL-CPG BINDING PROTEIN MB02.  
 GN MB02.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RC MEDLINE=98449942; PubMed=9774669;  
 RA Hendrich B., Bird A.;  
 RT "Identification and characterization of a family of mammalian methyl-  
 RT Cpg binding proteins.";  
 RL Mol. Cell. Biol. 18:6538-6547(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RC MEDLINE=99373255; PubMed=10441743;  
 RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;  
 RT "Genomic structure and chromosomal mapping of the murine and human  
 RT mbp1, mb02, mb03, and mb04 genes.";  
 RL Mamm. Genome 10:906-912(1999).  
 RL EMBL; AF072243; AAC66872.1; -;  
 DR EMBL; AF120986; AAD50372.1; -;  
 DR EMBL; AF120983; AAD50372.1; JOINED.  
 DR EMBL; AF120984; AAD50372.1; JOINED.  
 DR EMBL; AF120985; AAD50372.1; JOINED.  
 DR MGD; MGI:1333813; Mb02.  
 DR InterPro: IPR001739; MB0.  
 DR Pfam: PF01429; MB0.  
 DR SMART; SM00391; MB0; 1.  
 SO SEQUENCE 414 AA; 4354 MW; 9601D95E347BE53 CRC64;

Query Match	98.8%	Score 986	DB 11	Length 414
Best Local Similarity	98.5%	Pred. No. 2,76-75		
Matches 194	Conservative	2	Mismatches 1	Indels 0
			Gaps	0
OY	1	PSKIQKKKKRRLNDPLNKNKGRFDLTLLPIQOTASIFKQPYATVYTNHPSKRVASDPORM	60	
DB	218	PSKIQKKKKRRLNDPLNKNKGRFDLTLLPIQOTASIFKQPYATVYTNHPSKRVASDPORM	277	
OY	61	NEOPROLFMEKRLQGLSASDVTBOIITKTELPKRGQGVGPGSNDETLLSAVASALHTSSA	120	
DB	278	NEOPROLFMEKRLQGLSASDVTBOIITKTELPKRGQGVGPGSNDETLLSAVASALHTSSA	337	
OY	121	PITGVSAAEVKNPAWLVNTSOPLCAPFIVTDEDIRKQEEVYQGVRRKLEFALMAQILSR	180	
DB	338	PITGVSAAEVKNPAWLVNTSOPLCAPFIVTDEDIRKQEEVYQGVRRKLEFALMAQILSR	337	
OY	181	AAATPEENDIDMDSDEA	197	
DB	398	AAATPEYDIDMDSDEA	414	
RESULT	4			
O9PUM9				
ID	O9PUM9	PRELIMINARY	PRT	282 AA.
AC	O9PUM9			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	MEHLY-CPG BINDING PROTEIN MBD3.			
GN	MBD3			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipilidae;			
OC	Xenopodinae; Xenopus.			
NCBI	taxid=8335;			

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using SW model

Run on: July 24, 2002, 11:23:11 ; Search time 343.74 Seconds  
(without alignments)  
145,949 Million cell updates/sec

Title: PCT-US02-13994-27  
Perfect score: 1536  
Sequence: 1 HIKOKHFNERASRVVRYDV.....RRRALAGRGKENSPTAL 290

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*  
15: SP-virus:\*  
16: SP-bacteriap:\*  
17: SP-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1528	99.5	424	4	000312
2	1497.5	97.5	465	4	008085
3	1436	93.5	415	11	008605
4	1258.5	81.9	422	13	09YGM0
5	1125.5	73.3	465	4	09HBM9
6	1125.5	73.3	472	4	09Y2M6
7	1116	72.7	412	11	008606
8	984	64.1	316	4	09WT28
9	984	64.1	414	4	09HBM8
10	981	63.9	414	4	09HBM8
11	756.5	49.3	1090	5	09VG15
12	756.5	49.3	1142	5	09VG14
13	751.5	48.9	1150	5	094548
14	715	46.5	206	11	09D893
15	601	39.1	158	4	09WY89
16	596.5	38.8	761	5	022005

17	417	27.1	208	11	09CWM9	09CWM9 mus musculu
18	403	26.2	733	4	096J38	096J38 homo sapien
19	401	26.1	758	6	09BE29	09BE29 macaca fasc
20	400	26.0	382	4	016644	016644 homo sapien
21	399	26.0	733	11	09D2C0	09D2C0 mus musculu
22	388.5	25.3	456	4	09BYX9	09BYX9 homo sapien
23	388.5	25.3	473	11	054992	054992 mus musculu
24	382	24.9	471	4	060491	060491 homo sapien
25	379	24.8	735	4	096C05	096C05 homo sapien
26	379	24.7	737	13	09PRN1	09PRN1 mus musculu
27	377.5	24.6	490	10	042396	042396 xenopus lae
28	377	24.5	350	5	025108	025108 arabidopsis
29	376	24.5	493	5	015872	015872 hemilectrot
30	369	24.0	531	10	092SA2	092SA2 paramacium
31	369	24.0	531	10	094900	094900 arabidopsis
32	368	24.0	802	4	075582	075582 homo sapien
33	366	23.8	772	4	075676	075676 homo sapien
34	365.5	23.8	309	11	008763	008763 rattus norv
35	365.5	23.8	477	11	091WB2	091WB2 mus musculu
36	365.5	23.8	646	10	038870	038870 arabidopsis
37	365	23.8	460	4	095523	095523 homo sapien
38	365	23.8	476	4	096NM5	096NM5 homo sapien
39	361	23.5	481	4	09Y3J7	09Y3J7 homo sapien
40	361	23.5	554	10	049717	049717 arabidopsis
41	360.5	23.5	518	10	09AXA7	09AXA7 oryza sativ
42	360.5	23.5	520	10	09S361	09S361 arabidopsis
43	360	23.4	773	11	0922B9	0922B9 mus musculu
44	360	23.4	773	11	091X18	091X18 mus musculu
45	358.5	23.3	573	10	P93838	P93838 cucurbita p

## ALIGNMENTS

RESULT 1  
ID 000312 PRELIMINARY: PRT: 424 AA.  
AC 000312;  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MNKL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxId:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE:97299869; PubMed-9155018;  
RA Fukunaga R., Hunter T.;  
RT "MNKL, a new MAP kinase-activated protein kinase, isolated by a novel  
RT expression screening method for identifying protein kinase  
RT substrates.";  
RL EMBL J. 16:1921-1933(1997).  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL: AB000409; BAI19885.1; -;  
DR HSSP: Q63450; 1A06.  
DR InterPro: IPR000719; Euk\_Pkinase.  
DR InterPro: IPR002290; Ser\_thr\_Pkinase.  
DR Pfam: PF00069; Pkinase; 1.  
DR SMART: SM00220; S\_TKc; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
DR PROSITE: PS0011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Serine/threonine-protein kinase; Transferase.  
SQ  
SEQUENCE 424 AA: 47402 MW: DB39E49EC0BED990 CRC64:

Query Match 99.5% Score 1528; DB 4; Length 424;  
Best Local Similarity 99.3% Pred. No. 9.7e-136;  
Matches 288; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HIKOKHFNERASRVVRYDVAAALDLTKGIARHDKPENILCEPSKVSPIKICDFDL 60

```

DB 135 H10K0KHNEERASRYVDVAALDLFTHTKGIARHDLKPEINILCESPEKVSPIKICDFDL 194
OY 61 GSGKLNNSCPTITPTPELTTPCGSAEYMAPEVEVFTDOATFYDKRCDLMSIGVLYIML 120
DB 195 GSGKLNNSCPTITPTPELTTPCGSAEYMAPEVEVFTDOATFYDKRCDLMSIGVLYIML 254
OY 121 SGYPFVGHGADCGMDRGECVRCVONKLFESIOEGKTEFPDKNAMHISSEAKDLISKLL 180
DB 255 SGYPFVGHGADCGMDRGECVRCVONKLFESIOEGKTEFPDKNAMHISSEAKDLISKLL 314
OY 181 VDAKOKLSAAQVLOHPWVGOGAPENGLPTPOVLORNSTMDLTLLFAEAIALNRQISOH 240
DB 315 VDAKOKLSAAQVLOHPWVGOGAPENGLPTPOVLORNSTMDLTLLFAEAIALNRQISOH 374
OY 241 EENELAEPEALADGICSMKLSPPCKSRLARRRALLAOGGERSPTAL 290
DB 375 EENELAEPEALADGICSMKLSPPCKSRLARRRALLAOGGERSPTAL 424

```

```

RESULT 2
O9BUB5 PRELIMINARY: PRT: 465 AA.
AC O9BUB5:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SIMILAR TO MAP KINASE-INTERACTING SERINE/THREONINE KINASE 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA; CHORIOCARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
EMBL BC002755. AAR02755.1;
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002290; Ser_thr_PKinase.
DR InterPro: IPR001245; Tyr_PKinase.
DR Pfam: PF00069; Pkinase; 1.
DR SMART: SM00220; STKc; 1.
DR SMART: SM00219; TyrKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 465 AA; 51342 MW; CAE225C35DCB2B43 CRC64;

```

```

Query Match 97.5%; Score 1497.5; DB 4; Length 465;
Best Local Similarity 87.0%; Pred. No. 8; 2e-133;
Matches 288; Conservative 2; Mismatches 0; Indels 41; Gaps 1;

```

```

OY 1 H10K0KHNEERASRYVDVAALDLFTHTKGIARHDLKPEINILCESPEKVSPIKICDFDL 30
DB 135 H10K0KHNEERASRYVDVAALDLFTHTKGIARHDLKPEINILCESPEKVSPIKICDFDL 194
OY 31 -----GIAHRDLKPEINILCESPEKVSPIKICDFDLGSGKLNNSCPTITPTPELT 79
DB 195 PTSAQVAGTGIARHDLKPEINILCESPEKVSPIKICDFDLGSGKLNNSCPTITPTPELT 254
OY 80 TPCGSAEYMAPEVEVFTDOATFYDKRCDLMSIGVLYIMLSGYPFVGHGADCGMDRG 139
DB 255 TPCGSAEYMAPEVEVFTDOATFYDKRCDLMSIGVLYIMLSGYPFVGHGADCGMDRG 314
OY 140 EYCRVCOKNLFESIOEGKTEFPDKNAMHISSEAKDLISKLLVDAKOKLSAAQVLOHPWV 199
DB 315 EYCRVCOKNLFESIOEGKTEFPDKNAMHISSEAKDLISKLLVDAKOKLSAAQVLOHPWV 374
OY 200 OGQAPENGLPTPOVLORNSTMDLTLLFAEAIALNRQISOHRENEELAEPEALADGICSM 259

```

```

DB 375 OGQAPENGLPTPOVLORNSTMDLTLLFAEAIALNRQISOHRENEELAEPEALADGICSM 434
OY 260 KLSPPCKSRLARRRALLAOGGERSPTAL 290
DB 435 KLSPPCKSRLARRRALLAOGGERSPTAL 465

```

```

RESULT 3
O08605 PRELIMINARY: PRT: 415 AA.
AC O08605:
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MAP KINASE INTERACTING KINASE.
GN MNK1 OR MNK2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=97299868; PubMed=9155017;
RA Wasikiewicz A.J., Flynn A., Proud C.G., Cooper J.A.;
RT "Mitogen-activated protein kinases activate the serine/threonine
RT kinases MNK1 and MNK2."
RT EMBL J. 16:1909-1920(1997).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: Y11091; CAW1965.1;
DR HSBP: 063450; 1A06.
DR MGD: MGI:894316; Mknk1.
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002290; Ser_thr_PKinase.
DR Pfam: PF00069; Pkinase; 1.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 415 AA; 46544 MW; 9FA872CA611B2DAA CRC64;

```

```

Query Match 93.5%; Score 1436; DB 11; Length 415;
Best Local Similarity 94.1%; Pred. No. 4; 5e-127;
Matches 269; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 1 H10K0KHNEERASRYVDVAALDLFTHTKGIARHDLKPEINILCESPEKVSPIKICDFDL 60
DB 123 H10K0KHNEERASRYVDVAALDLFTHTKGIARHDLKPEINILCESPEKVSPIKICDFDL 182
OY 61 GSGKLNNSCPTITPTPELTTPCGSAEYMAPEVEVFTDOATFYDKRCDLMSIGVLYIML 120
DB 183 GSGKLNNSCPTITPTPELTTPCGSAEYMAPEVEVFTDOATFYDKRCDLMSIGVLYIML 242
OY 121 SGYPFVGHGADCGMDRGECVRCVONKLFESIOEGKTEFPDKNAMHISSEAKDLISKLL 180
DB 243 SGYPFVGHGADCGMDRGECVRCVONKLFESIOEGKTEFPDKNAMHISSEAKDLISKLL 302
OY 181 VDAKOKLSAAQVLOHPWVGOGAPENGLPTPOVLORNSTMDLTLLFAEAIALNRQISOH 240
DB 303 VDAKOKLSAAQVLOHPWVGOGAPENGLPTPOVLORNSTMDLTLLFAEAIALNRQISOH 362
OY 241 EENELAEPEALADGICSMKLSPPCKSRLARRRALLAOGGERSPTAL 286
DB 363 EENELAEPEALADGICSMKLSPPCKSRLARRRALLAOGGERSPTAL 408

```

```

RESULT 4
O9YGW0 PRELIMINARY: PRT: 422 AA.
ID O9YGW0:
AC O9YGW0:
DT 01-MAY-1999 (TREMblrel. 10, Created)

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 24, 2002, 11:10:27 ; Search time 346.34 Seconds

(Without alignments)  
287.675 Million cell updates/sec

Title: PCT-US02-13994-18

Perfect score: 4677  
Sequence: 1 EFLSKSKSEPPPGGLNHSLP.....AGIGSLRCAQGETEAEAM 897

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /Geneseq\_032802:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4669	99.8	897	AAV07092	Colon cancer assoc
2	4651	99.4	1141	AAV43008	Human ORFX ORF2772
3	3512	75.1	716	AAV49958	Human histone deac
4	2842	56.5	967	AAV49957	Human histone deac
5	2022	43.2	417	AAV93526	Human histone deac
6	1818	38.9	1008	AAV78891	Human polypeptide
7	1818	38.9	1020	AAV78891	Human polypeptide
8	1732	37.0	1428	AAV97033	Caspase 8-Interact
9	1594	34.1	614	AAV92700	Human protein sequ
10	1581.5	33.8	1235	AAV58493	Drosophila melanog
11	1539.5	33.3	574	AAV97032	Caspase 8-Interact

12	1515	32.4	541	21	AAV42845	Human ORFX ORF2609
13	1247	26.7	328	21	AAV45473	Human secreted pro
14	1090	23.3	327	21	AAV45474	Human secreted pro
15	945	20.2	191	22	AAV94542	Human protein sequ
16	948	19.2	276	22	AAV92577	Human protein sequ
17	732.5	15.7	726	22	AAV93056	Human protein sequ
18	723.5	15.5	883	22	AAV62141	Drosophila melanog
19	701.5	15.0	552	21	AAV51590	Arabidopsis thalia
20	697.5	14.9	499	21	AAV51591	Arabidopsis thalia
21	695.5	14.9	468	21	AAV51592	Arabidopsis thalia
22	693.5	14.8	552	21	AAV09530	Arabidopsis thalia
23	689.5	14.7	499	21	AAV09531	Arabidopsis thalia
24	687.5	14.7	468	21	AAV09532	Arabidopsis thalia
25	660.5	14.1	361	22	AAV23362	Arabidopsis thalia
26	641.5	13.7	817	22	AAV95640	Novel human enzyme
27	621	13.3	675	21	AAV42743	Human protein sequ
28	621	13.3	686	21	AAV42757	Human ORFX ORF2521
29	573.5	12.3	330	22	AAV02410	Human ORFX ORF2507
30	527	11.3	363	22	AAV02410	Human ORFX ORF2521
31	496	10.6	184	22	AAV94652	Human histone deac
32	451	9.6	170	22	AAV95880	Human histone deac
33	445	9.5	112	22	AAV95880	Human histone deac
34	445	9.5	164	21	AAV43314	Human histone deac
35	374.5	8.0	335	20	AAV97110	Human histone deac
36	371.5	7.9	398	21	AAV07246	Human histone deac
37	363	7.8	364	21	AAV07247	Human histone deac
38	363	7.5	377	21	AAV07247	Human histone deac
39	352	7.5	334	22	AAV67166	Human histone deac
40	303.5	6.5	433	22	AAV88919	Human histone deac
41	302	6.5	482	18	AAV29324	Human histone deac
42	300	6.4	482	22	AAV67164	Human histone deac
43	300	6.4	482	22	AAV67164	Human histone deac
44	300	6.4	482	22	AAV67164	Human histone deac
45	300	6.4	488	21	AAV56985	Human prostate can

#### ALIGNMENTS

RESULT 1	AAV07092	standard: Protein: 897 AA.
ID	AAV07092:	
AC	AAV07092:	
XX	02-JUL-1999 (first entry)	
XX	Colon cancer associated antigen precursor sequence.	
XX	Cancer associated antigen; diagnosis; research; treatment: human;	
XX	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
XX	prostate cancer.	
XX	Homo sapiens.	
XX	WO9904265-A2.	
XX	28-JAN-1999.	
XX	15-JUL-1998:	98WO-0514679.
XX	22-JUN-1998:	98US-0102322.
XX	17-JUL-1997:	97US-0896184.
XX	10-OCT-1997:	97US-0061599.
XX	10-OCT-1997:	97US-0061765.
XX	10-OCT-1997:	97US-0948705.
XX	11-OCT-1997:	97GB-0021697.
XX	(LUDWIG) LUDWIG INST CANCER RES.	
XX	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;	
XX	Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;	
XX	Tureci O;	

XX WP1: 1999-132448/11.  
 DR New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Dislosure: Page 652-654; 787pp: English.

XX The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.

SQ Sequence 897 AA:

Query Match 99.8%; Score 4669; DB 20; Length 897;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 896; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EFLSKSEKPTPGGLNLSLPOHPRKCGAHASLSDSSPPSGPPSTYKLP,PGPYAS 60  
 DB 1 EFLSKSEKPTPGGLNLSLPOHPRKCGAHASLSDSSPPSGPPSTYKLP,PGPYAS 60  
 QY 61 RDDPPLKRTASEPVLKRSRLKQVAERSSPLLRKDKGVISTRKRAVEITGAGPGAS 120  
 DB 61 RDDPPLKRTASEPVLKRSRLKQVAERSSPLLRKDKGVISTRKRAVEITGAGPGAS 120  
 QY 121 SVCSNAPGSPSPNSHSTIAENGFTGSVPNPTMELPOHALLDSSPNQFSLYTSPS 180  
 DB 121 SVCSNAPGSPSPNSHSTIAENGFTGSVPNPTMELPOHALLDSSPNQFSLYTSPS 180  
 QY 121 SVCSNAPGSPSPNSHSTIAENGFTGSVPNPTMELPOHALLDSSPNQFSLYTSPS 180  
 DB 121 SVCSNAPGSPSPNSHSTIAENGFTGSVPNPTMELPOHALLDSSPNQFSLYTSPS 180  
 QY 181 LPNLSLGLQATVTNNSHLTASPRLSLQDAERQALQSLRSGGTLTGFMSTSSIPGCLL 240  
 DB 181 LPNLSLGLQATVTNNSHLTASPRLSLQDAERQALQSLRSGGTLTGFMSTSSIPGCLL 240  
 QY 241 GVALEGGDSPHGHASLLQHVLLLEQARQOSTLAVPLHGSPLVMSGVATSMRTVGLP 300  
 DB 241 GVALEGGDSPHGHASLLQHVLLLEQARQOSTLAVPLHGSPLVMSGVATSMRTVGLP 300  
 QY 301 RHRPLSRTOSSPLFOSPOALQOLVMQOQHOOFLEROKOQOLQGLTLTKGELPPOPTTH 360  
 DB 301 RHRPLSRTOSSPLFOSPOALQOLVMQOQHOOFLEROKOQOLQGLTLTKGELPPOPTTH 360  
 QY 361 PEETFEELTEGOEVLGEGALTMPREGSTSESTOEDLEDEDEDEDEDEDEDEDEDEDE 420  
 DB 361 PEETFEELTEGOEVLGEGALTMPREGSTSESTOEDLEDEDEDEDEDEDEDEDEDEDE 420  
 QY 421 GESGAEGDPDEEPGAGYKRLFSDAOPLQYQAPLSLATVPHQALQRTOSSPAAPG 480  
 DB 421 GESGAEGDPDEEPGAGYKRLFSDAOPLQYQAPLSLATVPHQALQRTOSSPAAPG 480  
 QY 481 MKNPPDOYVVKLFTTSVYVOTFMKLRQCMGKRNHVRBAGKRIQISWRLQETGLSKCE 540  
 DB 481 MKNPPDOYVVKLFTTSVYVOTFMKLRQCMGKRNHVRBAGKRIQISWRLQETGLSKCE 540  
 QY 541 RIRGRKATLDEIQTVHSEYHTLTYGTSPLNRKQLDSKRLGPISONMTAVLPCGGIGVDS 600  
 DB 541 RIRGRKATLDEIQTVHSEYHTLTYGTSPLNRKQLDSKRLGPISONMTAVLPCGGIGVDS 600  
 QY 601 DTVMHMHSSSAVMAGCLLEAFKVAAGELNGFAIIRPGHAAESSTAMGCFERNV 660  
 DB 601 DTVMHMHSSSAVMAGCLLEAFKVAAGELNGFAIIRPGHAAESSTAMGCFERNV 660

DB 601 dtvmhmhsssaVMAGCLLEAFKVAAGELNGFAIIRPGHAAESSTAMGCFERNV 660  
 QY 661 AITAKLLOKLNKGVLYDMDIHHNGKTOQAFVNDPEVLISLHRYDNKNFPGSAGPE 720  
 DB 661 AITAKLLOKLNKGVLYDMDIHHNGKTOQAFVNDPEVLISLHRYDNKNFPGSAGPE 720  
 QY 721 EVGGGPGCVGVNVAVMTGVDPPIGDVEYLTAFTVNPVIAHEFSPDVLYVSGFPAVEG 780  
 DB 721 EVGGGPGCVGVNVAVMTGVDPPIGDVEYLTAFTVNPVIAHEFSPDVLYVSGFPAVEG 780  
 QY 781 HLSPLGCVSTAFPGHILTRLOMTLAGGVNVALBGGDULTRIDQASACVSLSTKID 840  
 DB 781 HLSPLGCVSTAFPGHILTRLOMTLAGGVNVALBGGDULTRIDQASACVSLSTKID 840  
 QY 841 PLDAVADQOPNINAVATLEKYLEIOSKHMSCVOKFAAGLGRSLRGACSTERAEM 897  
 DB 841 PLDAVADQOPNINAVATLEKYLEIOSKHMSCVOKFAAGLGRSLRGACSTERAEM 897

## RESULT 2

AAA43008  
 ID AAA43008 standard; Protein; 1141 AA.

AC AAA43008:

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2772 polypeptide sequence SEQ ID NO:5544.

KM Human; open reading frame; ORFX: detection; cytostatic; hepatotropic;  
 KM vulnery; antiproliferic; antiparkinsonian; nootropic; neuroprotective;  
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KM antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KM antidiabetic; gene therapy; cancer; proliferative disorder; hypertension;  
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KM cholesterol ester storage; systemic lupus erythematosus; infection;  
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KM thrombosis; contraceptive.

OS Homo sapiens.

PN MO200058473-A2.

PD 05-OCT-2000.

XX 31-MAR-2000; 2000MO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Leach M;

XX WP1: 2000-602362/37.

XX N-FSDB: AAC77217.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -

PS Claim 11: Page 4728-4731; 5507pp: English.  
 CC AAC74446 to AAC77606 encode the proteins given in AAA40237 to AAA43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OH protein - protein search, using sw model

Run on: July 24, 2002, 11:10:51 : Search time 346.34 seconds  
(without alignments)  
241.172 Million cell updates/sec

241.172 Million cell updates/sec

Title: PCT-US02-13994-21  
Perfect score: 3941  
Sequence: 1 RVKATLSERKIGDSCDKLP.....EASTWVQIFONTPTPL 752

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3941	100.0	752	20	AAV07108
2	2197	55.7	682	22	ABG3259
3	328.5	8.3	2515	22	ABG6447
4	273.5	6.9	777	22	AAU07898
5	268	6.8	928	22	AAU07880
6	191.5	4.9	1098	22	AAU03586
7	176	4.5	579	21	AAAB1958
8	166	4.2	3248	17	AAAB9795
9	165	4.2	2482	16	AAAB2836
10	165	4.2	2482	19	AAAB3996
11	160.5	4.1	633	22	AAAB3452

12	146.5	3.7	926	22	ABG62874
13	141	3.6	1019	22	AAU07892
14	139.5	3.5	1160	22	ABG17343
15	139	3.5	1543	22	ABG69213
16	138	3.5	944	22	ABG06345
17	136	3.5	1513	14	AAAB3253
18	135	3.4	355	21	AAAB2775
19	135	3.4	429	21	AAAB2774
20	135	3.4	496	21	AAAB2773
21	133.5	3.4	1788	22	AAAB0467
22	133	3.4	794	22	AAAB1978
23	133	3.4	933	21	AAAB6598
24	133	3.4	937	21	AAAB6597
25	133	3.4	985	21	AAAB6596
26	133	3.4	2025	22	AAU34207
27	133	3.4	2096	21	AAAB1592
28	133	3.4	3158	22	AAU37016
29	132.5	3.4	1743	19	AAAB6879
30	132.5	3.4	1786	18	AAAB24790
31	132	3.3	976	22	AAAB6581
32	132	3.3	1979	21	AAAB18171
33	131.5	3.3	1432	18	AAAB14517
34	131.5	3.3	2048	22	AAAB40027
35	130.5	3.3	2339	22	AAAB47278
36	130	3.3	3263	22	AAAB67210
37	129.5	3.3	864	22	AAAB39026
38	129	3.3	391	22	ABG20656
39	129	3.3	1049	22	ABG58717
40	128	3.2	1096	22	ABG62318
41	127.5	3.2	2271	22	ABG65616
42	127	3.2	745	22	AAAB07169
43	126.5	3.2	1144	21	AAAB6722
44	126.5	3.2	1144	21	AAAB5179
45	126.5	3.2	1703	21	AAAB36714

## ALIGNMENTS

RESULT 1	
AAV07108	AAV07108 standard; Protein: 752 AA.
XX	
AC	AAV07108:
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Colon cancer associated antigen precursor sequence.
XX	
KW	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer.
XX	
OS	Homo sapiens.
XX	
PN	W09904265-A2.
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998: 98W0-US14679.
XX	
PR	22-JUN-1998: 98US-0102322.
PR	17-JUL-1997: 97US-0896164.
PR	10-OCT-1997: 97US-0061599.
PR	10-OCT-1997: 97US-0061765.
PR	10-OCT-1997: 97US-0948705.
PR	11-OCT-1997: 97GB-0021697.
XX	
PA	(LUDWIG) LUDWIG INST CANCER RES.
XX	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI	Pfeundschuh M, Sahlin U, Scanlan MJ, Stockert E;
PI	Tureci O;

Drosophila melanog  
Polyprotein sequen  
Novel human diagn  
Drosophila melanog  
Novel human diagn  
p190 protein. Rat  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human polyprotein  
S. epidermidis ope  
Arabidopsis thalia  
Arabidopsis thalia  
Staphylococcus aur  
Human ORF ORF1356  
Staphylococcus aur  
H. pylori GHPD 175  
P. falciparum live  
Human SCP-1 mutain  
Plasmodium falcipa  
Human WRN gene pro  
Human polyprotein  
PN7771. Homo sapi  
Drosophila melanog  
Human polyprotein  
Novel human diagn  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Human HuFRG-1 pro  
S. cerevisiae DNA  
Cellulose synthase  
Arabidopsis thalia



XX WPI, 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -  
PT Isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers

XX Disclosure: Page 678-680; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.

XX Sequence 752 AA:

Query Match 100.0% Score 3941; DB 20; Length 752;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RVATLSERKICDSCDLPKCFEPKPTIMPGFTTVVSHINDLSFYVQIREDDEAE 60  
DB 1 RVKATLSERKICDSCDLPKCFEPKPTIMPGFTTVVSHINDLSFYVQIREDDEAE 60  
QY 61 ISHLSERLNSVTRPPTVYVGPPLQRCGMICAVPEEDNLMTRAIVIKQDQNDLSTVOFIDY 120  
DB 61 ISHLSERLNSVTRPPTVYVGPPLQRCGMICAVPEEDNLMTRAIVIKQDQNDLSTVOFIDY 120  
QY 121 GNVSVYVTHNKIGRDLVNAIIPGICINCSLOGFEVPNNKSKMMHYFQRTSEAAIRRE 180  
DB 121 GNVSVYVTHNKIGRDLVNAIIPGICINCSLOGFEVPNNKSKMMHYFQRTSEAAIRRE 180  
QY 181 FVKFQDRWEVLIADRHGIIADDMISRYALSRVASEQVELSTGVIKSASSKSVNSSDIDTSVF 240  
DB 181 FVKFQDRWEVLIADRHGIIADDMISRYALSRVASEQVELSTGVIKSASSKSVNSSDIDTSVF 240  
QY 241 LMMYNERKMIKAVATVIDSPETFCOPADTERKLCLEVEVQVAGQVADRRCICPPI 300  
DB 241 LMMYNERKMIKAVATVIDSPETFCOPADTERKLCLEVEVQVAGQVADRRCICPPI 300  
QY 301 GDFCIVRREDGHYRRALITINICEDYLSVRLVDFGNIEDCVPKALMAIPSELLSVPMQ 360  
DB 301 GDFCIVRREDGHYRRALITINICEDYLSVRLVDFGNIEDCVPKALMAIPSELLSVPMQ 360  
QY 361 AFPCCLSGNINISBGLCSQSGNDFYEITTEVLEITLIRROYCDPLAIVDLKSKGS 420  
DB 361 AFPCCLSGNINISBGLCSQSGNDFYEITTEVLEITLIRROYCDPLAIVDLKSKGS 420  
QY 421 INKMKRYSKGTIKSALPENIDSEIKQTLGSYNLDGLKKSINKAONKTYEEOOTDL 480  
DB 421 INKMKRYSKGTIKSALPENIDSEIKQTLGSYNLDGLKKSINKAONKTYEEOOTDL 480  
QY 481 AETTEDVNIIGTKPSNRPDPTDNICGFFNPCKKIDTDELBGELECHLVNKAFFDK 540  
DB 481 AETTEDVNIIGTKPSNRPDPTDNICGFFNPCKKIDTDELBGELECHLVNKAFFDK 540  
QY 541 YLTGNTLLPHANETKELIENSLFVPLSPDSEKFELESTELONSLVVOEFGELS 600  
DB 541 YLTGNTLLPHANETKELIENSLFVPLSPDSEKFELESTELONSLVVOEFGELS 600  
QY 601 PVPNPVLSQECVTKAMELFTQLPLSCGAKEKOPELLPALQPLDDKMPULSLGVSOK 660  
DB 601 PVPNPVLSQECVTKAMELFTQLPLSCGAKEKOPELLPALQPLDDKMPULSLGVSOK 660

DB 601 PVPNPVLSQECVTKAMELFTQLPLSCGAKEKOPELLPALQPLDDKMPULSLGVSOK 660  
QY 661 AOEKCTEDMRKSCVDFDORRMSLHAGADCPKTONENICEEFVEYKRRDAISA 720  
DB 661 AOEKCTEDMRKSCVDFDORRMSLHAGADCPKTONENICEEFVEYKRRDAISA 720  
QY 721 LMPESLRKKAAMEASTIMYQIIFONTPTPTL 752  
DB 721 LMPESLRKKAAMEASTIMYQIIFONTPTPTL 752

# RESULT 2

ABG23259 standard; Protein: 682 AA.  
ABG23259:  
18-FEB-2002 (first entry)  
Novel human diagnostic protein #23250.  
Human: Chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
OS WO200175067-A2.  
FN 11-OCT-2001.  
PD 30-MAR-2001; 2001WO-US08631.  
PF 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
PI WPI: 2001-639362/73.  
DR N-PSDB: AAS87446.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 53618; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/pmid/ished\_pct\_sequences.  
XX Sequence 682 AA:

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using SW model

Run on: July 24, 2002, 11:11:01 ; Search time 346.34 Seconds

(without alignments)  
63.179 Million cell updates/sec

Title: PCT-US02-13994-23

Perfect score: 1 PSLKLNKQRLRNPLNOK.....LSNADTEEMDMSGDEA 197

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: A_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	998	100.0	200	20	AAV07107
2	998	100.0	263	19	AAW74980
3	998	100.0	281	20	AAV73829
4	998	100.0	281	20	AAV48439
5	998	100.0	411	20	AAV14197
6	998	100.0	411	22	AAAB9915
7	998	100.0	411	22	AAAG64314
8	998	100.0	411	22	AAAG64844
9	985	98.8	414	20	AAV14199
10	748	74.9	223	22	AAAB92997
11	748	74.9	291	20	AAV14198

12	739.5	74.1	285	20	AAV14200
13	739	74.0	218	22	AAAB93733
14	340.5	34.1	226	22	AAAB63747
15	340.5	34.1	314	22	AAAB65035
16	107.5	10.8	1038	22	AAAB71916
17	101	10.1	248	22	AAAB71902
18	95.5	9.6	332	22	AAAB61802
19	95.5	9.6	1914	18	AAW24800
20	95.5	9.6	1312	20	AAAB3807
21	95.5	9.6	1312	20	AAV33495
22	91	9.1	1048	21	AAV68823
23	89.5	9.0	507	21	AAAB43044
24	89	8.9	214	21	AAAB3873
25	89	8.9	1203	22	AAW79264
26	88.5	8.9	568	22	AAAB63059
27	88.5	8.9	1313	19	AAW60213
28	88	8.8	212	21	AAAG25210
29	88	8.8	959	21	AAV77944
30	87.5	8.8	706	22	AAAB94174
31	87.5	8.8	1406	22	AAU32728
32	87	8.7	284	22	AAAB60998
33	87	8.7	408	22	AAAG12899
34	87	8.7	732	22	AAAB94010
35	87	8.7	970	21	AAAB41925
36	86.5	8.7	596	22	AAAB69163
37	86.5	8.7	5034	22	AAAG82935
38	86	8.6	521	21	AAAG31483
39	86	8.6	337	21	AAAG31482
40	86	8.6	540	21	AAAG31481
41	86	8.6	2665	22	AAAB28314
42	86	8.6	2665	22	AAAB33490
43	86	8.6	2665	22	AAAB18950
44	86	8.6	2665	22	AAW54270
45	86	8.6	2665	22	AAAB66655

#### ALIGNMENTS

#### RESULT 1

AAV07107 standard; Protein: 200 AA.

AAV07107:

02-JUL-1999 (first entry)

Colon cancer associated antigen precursor sequence.

Cancer associated antigen; diagnosis; research; treatment; human;

breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

prostate cancer.

XX Homo sapiens.

XX W09904265-A2.

XX 28-JAN-1999.

XX 15-JUL-1998: 98MO-US14679.

XX 22-JUN-1998: 98US-0102322.

XX 17-JUL-1997: 97US-0896164.

XX 10-OCT-1997: 97US-0061599.

XX 10-OCT-1997: 97US-0061765.

XX 11-OCT-1997: 97US-0948705.

XX (LUDWIG-) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Odata Y, Old LJ;

XX Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;

XX Tureci O;

Mouse DNA demethyl  
Human protein sequ  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Novel human diagno  
Drosophila melanog  
Sphinocebellar at  
Human ataxin-2. H  
Human SCA2 protein  
Amino acid sequenc  
Human ORF ORF2808  
Arabidopsis thalia  
Human protein S80  
Drosophila melanog  
Sphinocebellar at  
Arabidopsis thalia  
A. thaliana enviro  
Human protein sequ  
Novel human secret  
Drosophila melanog  
Novel human diagno  
Human protein sequ  
Human ORF ORF1689  
Drosophila melanog  
S. epidermidis ope  
Arabidopsis thalia  
Arabidopsis thalia  
Human peptide 1965  
Peptide 1996 encod  
Human brain expres  
Human Bone marrow

XX WPI; 1999-132448/11.  
DR

PT New isolated cancer associated nucleic acids and polypeptides -  
PT Isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers

PS Disclosure; Page 677; 787pp; English.

The invention relates to a method for diagnosing a disorder characterised by expression of a human cancer associated antigen precursor coded for by a nucleic acid molecule (NAM). The method comprises: (a) contacting a biological sample isolated from a subject with an agent that specifically binds to the NAM, an expression product or a fragment of an expression product complexed with an HLA molecule; and (b) determining the interaction between the agent and the NAM or the expression product as a determination of the disorder. The products and methods can be used in the diagnosis, monitoring, research, or treatment of conditions characterised by the expression of various cancer associated antigens. The invention provides nucleic acid sequences and encoded polypeptides which are cancer associated antigen precursors expressed in human breast cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and lung cancer.

**SQ Sequence 200 AA;**

Query Match	100.0%	Score 998;	DB 20;	Length 200;
Best Local Similarity	100.0%	Pred. No. 7.4e-91;		
Matches 197; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY 1 PSKLOKNKQRLRNDPLNQNKKGKPDINTLPIRQTASIFKQPVTKVTNHPSNKVKSDPQRM 60

Db 4 psklqknkqrlrnoplnqngkypdlnltlplrtqtaslfkqpvtkvtnhpsnkvksdpqrm 63

01 NEUPKOLEMEKRLQGSASDVTEQI IKIMELPKGLQGVGPGSNDETILSAVASALHTSSA 120

04 "eqpyqytwenrygtsasvceqywnuiepygyqgyvgypsndecr.18aVasSaInCtSsa 123

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible][illegible]

RESULT	2
AAW74980	
ID	AAW74980 standard; Protein; 263 AA

AC LAW74980;  
XX

DT 25-JAN-1999 (first entry)  
XX

Human secreted protein encoded by gene 106 clone HT3AM65.

human; secreted protein; testis; tumour; foetal brain tissue; fusion protein; cancer; central nervous system; seizure;

XX

XXXXXX

FT Misc-differer

XX  
PN  
W0983844B-13

XX  
PD 11-SEP-1998

06-MAR-1998; 98WO-US04493

1

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 24, 2002, 11:11:04 ; Search time 346.34 Seconds  
(Without alignments)  
174.786 Million cell updates/sec

Title: PCT-UB02-13994-25  
Perfect score: 2689  
Sequence: 1 METPSGRATRSQAGASSTP.....LRTALINSTGEVAMKLVLR 545

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_032802.\*

1:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT.*
21:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.*
22:	/SIDS1/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2689	100.0	572	20	AAV07079
2	2689	100.0	572	22	AAAB90757
3	1980	73.6	515	19	AAAS4236
4	1980	73.6	515	22	AAAB90770
5	1346.5	50.1	777	22	AAAB9238
6	1314	48.9	831	22	AAAB0222
7	979	36.4	622	22	ABAB62816
8	937.5	34.9	621	22	ABAB64883
9	575	21.4	120	20	AAV29554
10	575	21.4	120	21	AAAB4477
11	575	21.4	120	22	AAE13819
					Human Lung tumour-
					Human Lung tumour-
					Human Lung tumour-

12	454.5	16.9	466	22	AAAB6349	Murine vimentin.
13	449.5	16.7	465	19	AAAB9351	Vimentin, Homo sa
14	449.5	16.7	466	21	AAAB2935	Human poliovirus-d
15	449.5	16.7	466	21	AAV92335	Human vimentin. H
16	449.5	16.7	466	22	AAAB6348	Human vimentin. H
17	445.5	16.6	433	19	AAV20975	Human gliol fibril
18	445	16.5	543	22	AAV93466	Human polypeptide,
19	416	15.5	513	22	ABG08133	Novel human diagno
20	406	15.1	498	21	AAAB08480	Amino acid sequenc
21	397	14.8	430	22	AAAB90795	Human shear stress
22	397	14.8	447	22	ABG16550	Novel human diagno
23	397	14.8	452	22	AAAG74328	Human colon cancer
24	391.5	14.6	551	21	AAV52398	Human keratin KERT
25	384	14.3	1026	22	AAV78825	Human protein SBO
26	382	14.2	1033	22	AAV78809	Human protein SBO
27	366.5	13.6	508	22	AAV78373	Human protein SBO
28	365.5	13.6	495	22	AAV79357	Human protein SBO
29	353	13.1	73	22	ABBA44305	Peptide #11811 enc
30	353	13.1	73	22	ABBA27167	Human brain expres
31	353	13.1	73	22	AAV65342	Human bone marrow
32	353	13.1	73	22	AAV78037	Peptide #12295 enc
33	353	13.1	73	22	AAV65342	Human bone marrow
34	353	13.1	73	22	AAV23928	Peptide #12295 enc
35	352.5	13.1	433	21	AAAB8258	Breast and ovarian
36	351.5	13.1	546	21	AAV52397	Human keratin KERT
37	349	13.0	364	22	ABG27717	Novel human diagno
38	347.5	12.9	456	22	ABG15224	Novel human diagno
39	346	12.9	477	22	AAV006112	Novel human polype
40	342	12.7	469	18	AAV23820	Human sarcolectin.
41	342	12.7	469	21	AAV69289	Amino acid sequenc
42	341.5	12.7	486	22	ABG14794	Novel human diagno
43	341.5	12.7	505	22	AAV33352	Human keratin enco
44	341.5	12.7	505	22	AAV33354	Human breast cance
45	341.5	12.7	558	22	ABG23688	Novel human diagno

## ALIGNMENTS

RESULT	1
AAV07079	standard; Protein: 572 AA.
XX	
XX	AAV07079:
AC	02-JUL-1999 (first entry)
XX	
XX	Renal cancer associated antigen precursor sequence.
XX	
XX	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KM	prostate cancer.
XX	
XX	Homo sapiens.
OS	
PN	WC0904265-A2.
XX	
PD	28-JAN-1999.
XX	
XX	15-JUL-1998; 98WO-0514679.
PF	
XX	
PR	22-JUN-1998; 98US-0102322.
PR	17-JUL-1997; 97US-0896164.
PR	10-OCT-1997; 97US-0061599.
PR	10-OCT-1997; 97US-0061765.
PR	10-OCT-1997; 97US-0948705.
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDWIG) LUDWIG INST CANCER RES.
XX	
PI	Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
PI	Pfreundschuh M, Sahlin U, Scanlan MJ, Stockert E;
PI	Tureci O;

XX  
DR WPI: 1999-132448/11.

PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers

PS Disclosure: Page 485-387; 787pp; English.

XX  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.

SO Sequence 572 AA:

Query Match 100.0%; Score 2689; DB 20; Length 572;  
Best Local Similarity 100.0%; Pred. No. 1.4e-174;  
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METTSORRATRSQOASSTPLSPTRITRIQEKEDLOELMDRLAVYIDVRSLTENGLR 60  
DB 1 metpsqratrsgqasstplsptritrirgkedlqelndrlavldvrsltenaglr 60  
QY 61 LRITSEEVVSREVSIGIKAAVEAEICDARTIDSVAKERARQLQLSVRREFELKARN 120  
DB 61 lrftseevvsrevsgikaayeaeldartkldsvakerarqlqlsvrrefelkarn 120  
QY 121 TKKEDDILAAARLKDEALLNSFEALSTALSEKRTLGEHLDRGVAKLEALGEAK 180  
DB 121 tkkeddilaagarlkdeallnskaalstalsekrtlgehlhdirgvakeaalgeak 180  
QY 181 KOLDEMLRVDAENRLQTNKEELDFQKNITSEELRETRRRHETRLVEIDNCKOREFSR 240  
DB 181 koldemlrvdaenrlqtnkeeldfqnityselretrrrhetrlveidnckorefsr 240  
QY 241 LADALQELRAOHEOVQYKKELEKTSARLDNAROSAEBSNSLVGAHEELQOSRIRID 300  
DB 241 ladalqelraqhedvqykkelekyakldnarqsaenmslvgaheelqgsrirlid 300  
QY 301 STSAOLSOLOKOLAKEAKLRDLSDSLAREDTSRRLAEKEREMAEKRAMQOOLDEYQ 360  
DB 301 stsaolsolokolaekaklrldslaredtssrllaekeeremaeqramoooldeyq 360  
QY 361 ELUDTKIALDMHIAHYRRLGCEERRLRLSPPTSORBSRASHSSORTGGSVTKKRX 420  
DB 361 ellidtkialdmeihayrlllegeerlrllspptsgsrashssqg99svtkrkx 420  
QY 421 LESTESRSFSOHARTSGVAVEVEDEGKFEVRLRNSNEDQSMGMWQIKRONGDPLLT 480  
DB 421 lestesrsfsghartsgvaveevedegkfvrlrnksnedsqgmwqikrqgdpllt 480  
QY 481 YRRPPFTLAKAGOVTTAAAGATHSPPTDLVYKANTCGCNSRLTALINSTGEVAM 540  
DB 481 yrrppftlkagovttaaagathspptdlvvykagntvqcgnsrltalinstgevam 540  
QY 541 RKLVR 545  
DB 541 rklvr 545

RESULT 2

AAB90757  
ID AAB90757 standard; Protein: 572 AA.

AC AAB90757;  
XX  
AC 15-JUN-2001 (first entry)  
DT

DE Human shear stress-response protein SPQ ID NO: 14.

KW Human: shear stress-response protein; vascular disease;  
KM arteriosclerosis.

OS Homo sapiens.

PN W0200125427-A1.

PD 12-APR-2001.

PE 02-OCT-2000; 2000WO-JP06840.

PR 01-OCT-1999; 99JP-0280976.

PA (RYOM) KYOMA HAKKO KOGYO KK.

PI (NOJ1) NOJIMA H.

PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;  
PI Kuga T, Sekine S, Nakamura Y, Sugano S;

DR WPI: 2001-266308/27.

DR N-PSDB; AAH02880.

PT DNA sequences, proteins encoded by them and antibodies against them  
PT useful in diagnosis and treatment of vascular disease caused by

PT arteriosclerosis -

PS Claim 60; Page 192-195; 678pp; Japanese.

CC The present invention provides the protein and coding sequences of a  
CC number of human shear stress response proteins. These are useful in the  
CC diagnosis, treatment and screening of vascular diseases caused by  
CC arteriosclerosis, including heart failure, post-PTCA restenosis and  
CC hypertension.

SO Sequence 572 AA:

Query Match 100.0%; Score 2689; DB 22; Length 572;  
Best Local Similarity 100.0%; Pred. No. 1.4e-174;  
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METTSORRATRSQOASSTPLSPTRITRIQEKEDLOELMDRLAVYIDVRSLTENGLR 60  
DB 1 metpsqratrsgqasstplsptritrirgkedlqelndrlavldvrsltenaglr 60  
QY 61 LRITSEEVVSREVSIGIKAAVEAEICDARTIDSVAKERARQLQLSVRREFELKARN 120  
DB 61 lrftseevvsrevsgikaayeaeldartkldsvakerarqlqlsvrrefelkarn 120  
QY 121 TKKEDDILAAARLKDEALLNSFEALSTALSEKRTLGEHLDRGVAKLEALGEAK 180  
DB 121 tkkeddilaagarlkdeallnskaalstalsekrtlgehlhdirgvakeaalgeak 180  
QY 181 KOLDEMLRVDAENRLQTNKEELDFQKNITSEELRETRRRHETRLVEIDNCKOREFSR 240  
DB 181 koldemlrvdaenrlqtnkeeldfqnityselretrrrhetrlveidnckorefsr 240  
QY 241 LADALQELRAOHEOVQYKKELEKTSARLDNAROSAEBSNSLVGAHEELQOSRIRID 300  
DB 241 ladalqelraqhedvqykkelekyakldnarqsaenmslvgaheelqgsrirlid 300  
QY 301 STSAOLSOLOKOLAKEAKLRDLSDSLAREDTSRRLAEKEREMAEKRAMQOOLDEYQ 360  
DB 301 stsaolsolokolaekaklrldslaredtssrllaekeeremaeqramoooldeyq 360

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 24, 2002, 11:10:35 : Search time 346.34 seconds

(without alignments)  
285,430 Million cell updates/sec

Title: PCT-US02-13994-19

Perfect score: 4413  
Sequence: 1 MFYMDDELKSESVFROLN.....RQDHQDKKDGITPAQLVNY 890

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4016	91.0	1068	AAV59273	Mouse huntingtin-1
2	3296	74.7	676	AAV59271	Human huntingtin-1
3	2014.5	45.6	1090	AAV59270	Human huntingtin-1
4	2009.5	45.5	914	AAV18030	Huntingtin interact
5	2009.5	45.5	914	AAV59269	Human huntingtin-1
6	1640.5	37.2	756	AAV59272	Human huntingtin-1
7	1400	31.7	282	AAV59272	Human protein sequ
8	896.5	20.3	1087	ABR65358	Drosophila melanog
9	896.5	20.3	1087	ABR65359	Drosophila melanog
10	718.5	16.3	388	AAV18039	Huntingtin interact
11	705.5	16.0	386	AAV59268	Human huntingtin-1

12	375.5	8.5	2541	21	AAV41087
13	358	8.1	1134	22	AAV78810
14	358	8.1	1144	22	AAV39293
15	358	8.1	1165	22	AAV39312
16	353	8.0	2836	22	ABR62719
17	348	7.9	1177	22	AAV79794
18	348	7.9	1177	22	AAV41079
19	348	7.9	1177	22	AAV41098
20	332	7.5	2101	21	AAV49936
21	326	7.4	2192	18	AAV21732
22	326	7.4	2272	18	AAV21731
23	325	7.4	1154	22	AAV32042
24	325	7.4	1960	22	AAV78854
25	325	7.4	2143	22	ABG01716
26	324	7.3	2117	22	AAV32040
27	322	7.3	1690	22	ABR61144
28	322	7.3	1690	22	ABR61173
29	322	7.3	2101	15	AAV41173
30	322	7.3	2101	22	AAV65799
31	322	7.3	2115	21	AAV49937
32	321	7.3	2207	22	AAV32041
33	320	7.3	118	22	AAV94649
34	320	7.3	118	22	AAV18821
35	318	7.2	1886	19	AAV54241
36	318	7.2	2246	22	ABG05850
37	316.5	7.2	1963	22	AAV79838
38	312.5	7.1	2442	21	AAV77575
39	311.5	7.1	1988	22	AAV40999
40	311.5	7.1	1988	22	AAV41000
41	305.5	6.9	1083	22	ABG20032
42	305.5	6.9	1851	22	ABG01723
43	302	6.8	1879	22	AAV25750
44	299.5	6.8	1499	22	ABR59948
45	298.5	6.8	2056	22	ABR59344

## ALIGNMENTS

RESULT 1  
AAV59273 standard: Protein: 1068 AA.  
XX  
AC AAV59273:  
XX  
DT 17-APR-2000 (first entry)  
XX  
XX Mouse huntingtin-Interacting protein (mHt1p1a).  
DE  
XX Huntingtin Interacting Protein; HRP; death effector domain; DED; human;  
XX Apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;  
KW Huntingtin's disease; nootropic; anticonvulsant; cytostatic; mouse.  
XX  
XX Mus sp.  
OS  
XX  
XX WO9960986-A2.  
XX  
XX 02-DEC-1999.  
XX  
XX 27-MAY-1999: 99WO-US11743.  
XX  
XX 27-MAY-1998: 98US-0085199.  
XX  
XX (UYBR-) UNIV BRITISH COLUMBIA.  
PA (MERRI) MERCK PROSSER CANADA INC.  
XX  
XX Kalkman M, Hayden MR, Hackam A, Chopra V, Nicholson DW:  
PI Vallatincourt JP, Rasper DM:  
XX  
XX WPI: 2000-097055/08.  
XX N-PSDB: AA258749.  
XX  
XX Novel proteins useful for treating Huntington's disease by gene therapy

PT techniques, and cancers .

XX Example 11: Page 73-77; 91pp; English.

CC The invention relates to Huntingtin Interacting Protein (HIP), that  
CC includes a death effector domain (DED), suggesting apoptotic function.  
CC Proteins with DED (referred to as HIP-apoptosis modulating proteins) are  
CC useful for inducing apoptotic death in cells. The HIP is a logical target  
CC for therapy in Huntington's disease since it has a differential apoptotic  
CC activity, modulated by interaction with Huntingtin having normal and  
CC expanded repeats. HIP is also used as a therapeutic agent to introduce  
CC apoptosis in cancer cells. Increasing expression of normal (non-expanded)  
CC Huntingtin or the HIP-apoptotic modulator-binding portion, a modified  
CC HIP-apoptotic modulator in which the DED has been deleted, is helpful for  
CC treating Huntington's disease. The present sequence represents the mouse  
CC HIPa.

XX Sequence 1068 AA:

Query Match 91.0%; Score 4016; DB 21; Length 1068;

Best Local Similarity 90.0%; Pred. No. 1,3e-254;

Matches 801; Conservative 44; Mismatches 45; Indels 0; Gaps 0;

OY 1 MFYMDCELKLSVFRQNTAIVSOMSGCRLAPLIQVITQDCSHLYHTVNLKFLH 60  
DB 179 mfdymdelklsesvfrqntalavsgmasgqcrilapliqvldeshlyhtvnlkflh 238  
OY 61 SCLEADITLQGRHDFHOFSLNFRFRASDMYIFRLIQLIRLDEGPNFLRASALAEH 120  
DB 239 sclpaditlqgrhdfhofslnfrfrasdmlyifrlilqlirldegpnlrasalae 298  
OY 121 IKPVVIFPEEAPDEDEPENLEISTGPAPGEVYVADLDFQFGPNKSVYDDDDLOIES 180  
DB 299 ikpvvifpeeapeeepenleisatpgepyvadvldqfipgnsmddldiqen 358  
OY 181 LKREYMLRSLEKIKLEAORTIAOLKSOVNALEGEEROKOKALVNDQRLHQLAQ 240  
DB 359 lkrevellreleiklmeagrylsqikgvnglealeeqkqkqlvndeqrlhelaq 418  
OY 241 LRAALGERSOGILREAEERKASATFARYNKLKEKSELVHYHALLKRNADTKOYLTV 300  
DB 419 lkalqlagarnqglreecerksatearysklkehseilnhaelirhndatkkqltvc 478  
OY 301 QQSEEFYARVBOALFOVOYKRESSELKLEKSDOLEKIKRELEAKAGELARAQALSH 360  
DB 479 qsqseefaryvboalfvovoykresesleklesdolekikrelaagelaraqalsht 538  
OY 361 EGSSELSSRLDTLSAKRDALSGAVRQREADLLAQSIVREFEALSDEQORSSOBGEL 420  
DB 539 egsgselssrltdltsakrdalsgavrqrealllaqslvrekelaadegqqrssqekyel 598  
OY 421 QGRLAERESQEGRLRLDEQFAVIRGAALAEAGILDAVSKLDDPLRLHCTSSPYLV 480  
DB 599 rgqlaereseqgrlrldeqlavirrsaaaeaalqdvavkliddphrlhctsspylv 658  
OY 481 SRRAELDAVSTEEGHAQYLTSLADASALVVALTRFSLADRTINGASHSALPDPRA 540  
DB 659 srqaaladvsteeqghatqylasasdasalvaltrfsladrtlvngaaahlapdp 718  
OY 541 DRLIDTCRBCGAPALFLMGQLDQOALRHMQASLVRTPLQGLTQIGRELKPKSLDVQEE 600  
DB 719 drlidtrcrgaralelvqqlqdvltlrraqplmraplqqlqgqqlkpksldivqee 778  
OY 601 LGAVVDEKMATSAIEDAVRTEEDMNQARHASSGVLEFVNERLINSCTDLMKATILV 660  
DB 779 lgvavdekmatseaiedavrteddmnqarhassgvlefvnerlinsctdlnkatilv 838  
OY 661 TTSTSLQKREIVESGKAATDOEFYANSRMTGLISAKAVMGATOLVLEADRVYLTG 720  
DB 839 mtstslqkreivesgkaatqefyknsvrtglisakavmgatqlvleadvlyltmg 898  
OY 721 KYEELIVCSHRILASTAQVLAASKVYANRHSPLSLDECSRIVNERANVYASTKSGOE 780

DB 899 kyeelivcsheiaastaglvaaakvanknspshisldcearlvneraanvaastksgoe 958

OY 791 QIEDRTMDPSGLSLKLNKQENETQVRYLEKTLAEARNLSELKROHYVLGASGSP 840

DB 959 qiedrtdmdfsgllslklkqemetgvrylelekleerlvrlgelrtqhyvlagmgtp 1018

OY 841 GEFAVIRPSTAPSVTTTKRPPLAQKPSVAPRODDHQLDKKDCITPAQLVNY 890

DB 1019 geefavirpstapsvtttkrpplaqkpsvaproddhqldkkdqitpaqlvny 1068

RESULT 2

AAV59271

ID AAV59271 standard; Protein; 676 AA.

XX AAV59271:

DT 17-APR-2000 (first entry)

DE Human huntingtin-interacting protein (HIP1).

KM Huntingtin Interacting Protein: HIP; death effector domain; DED; human;

KM apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;

KM Huntington's disease; neurotropic; anticonvulsant; cytostatic;

XX Chromosome 7q11.23.

OS Homo sapiens.

PM MO9960986-A2.

PD 02-DEC-1999.

PF 27-MAY-1999; 99MO-US11743.

PR 27-MAY-1998; 98US-0085199.

PA (UYBR-) UNIV BRITISH COLUMBIA.

PA (MERI) MERCK FROST CANADA INC.

PI Kalchman M, Hayden MR, Hackam A, Chopra V, Nicholson DM;

PI Vallatincourt JP, Rasper DM;

DR MPI: 2000-097055/08.

DR N-PSDB; AAZ58747.

PT Novel proteins useful for treating Huntington's disease by gene therapy

PT techniques, and cancers .

Claim 3; Page 63-66; 91pp; English.

CC The invention relates to Huntingtin Interacting Protein (HIP), that  
CC includes a death effector domain (DED), suggesting apoptotic function.

CC Proteins with DED (referred to as HIP-apoptosis modulating proteins) are  
CC useful for inducing apoptotic death in cells. The HIP is a logical target  
CC for therapy in Huntington's disease since it has a differential apoptotic  
CC activity, modulated by interaction with Huntingtin having normal and  
CC expanded repeats. HIP is also used as a therapeutic agent to introduce  
CC apoptosis in cancer cells. Increasing expression of normal (non-expanded)  
CC Huntingtin or the HIP-apoptotic modulator-binding portion, a modified  
CC HIP-apoptotic modulator in which the DED has been deleted, is helpful for  
CC treating Huntington's disease. The present sequence represents the human  
CC HIP1.

XX Sequence 676 AA:

Query Match 74.7%; Score 3296; DB 21; Length 676;

Best Local Similarity 99.7%; Pred. No. 1.1e-207;

Matches 674; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 215 GELEQRKOKALVNDQRLHQLRAQLDEGRSGGLREAEKRASTFARYKLTKE 274

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 11:24:45 : Search time 95.25 Seconds

(without alignments)  
73.171 Million cell updates/sec

Title: PCT-US02-13994-30

Perfect score: 959  
Sequence: 1 MQAECRGSTGSDGADGPG.....WITGCFPLVFLAQPSSGRR 180

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	959	100.0	180	1	CTAG_HUMAN
2	595	62.0	210	1	LAGI_HUMAN
3	174.5	18.2	627	1	HYDL_STRCO
4	155.5	16.2	641	1	EBN1_EBV
5	152.5	15.9	779	1	CA11_BOVIN
6	151.5	15.8	423	1	BR3A_HUMAN
7	151.5	15.8	1464	1	CA11_HUMAN
8	151	15.7	1901	1	YZ08_MYCTU
9	148.5	15.5	252	1	GRP1_PRAVU
10	148	15.4	1355	1	CA21_RANCA
11	148	15.4	1356	1	CA21_ONCMY
12	147.5	15.4	1453	1	CA11_CHICK
13	147.5	15.4	1460	1	CA11_CANFA
14	146.5	15.3	1806	1	CA1B_HUMAN
15	146	15.2	338	1	GRP_ARATH
16	145.5	15.2	482	1	CA1B_RAT
17	144.5	15.1	301	1	CC02_CAEEL
18	144.5	15.1	1147	1	MY58_ACACA
19	143.5	15.0	421	1	BR3A_MOUSE
20	143.5	15.0	1049	1	CA13_BOVIN
21	143.5	15.0	1262	1	CA13_CHICK
22	143	14.9	323	1	Y034_CAEEL
23	142.5	14.9	1466	1	CA13_HUMAN
24	142	14.8	1372	1	CA21_MOUSE
25	141.5	14.8	1366	1	CA21_HUMAN
26	141	14.7	1418	1	CA12_HUMAN
27	140.5	14.7	671	1	CA11_RAT
28	140.5	14.7	1453	1	CA11_MOUSE
29	140.5	14.6	1603	1	CA1F_HUMAN
30	140	14.6	1459	1	CA1E_MOUSE
31	139.5	14.5	1372	1	CA21_RAT
32	139	14.5	463	1	SANT_PLAIV
33	139	14.5	463	1	YA68_MYCTU

## ALIGNMENTS

RESULT ID	1	CTAG_HUMAN	STANDARD:	PRT:	180 AA.
AC	P78358				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Autoimmunogenic cancer/testis antigen NT-ESO-1.				
GN	CTAG.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_Taxid:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-97203161; PubMed-9050879;				
RA	Chen Y.-T., Scanlan M.J., Sahin U., Tuercel O., Gure A.O., Tsang S.,				
RA	Williamson B., Stockert E., Pfundschuh M., Old L.J.,				
RT	"A testicular antigen aberrantly expressed in human cancers detected				
RT	by autologous antibody screening."				
RL	Proc. Natl. Acad. Sci. U.S.A. 94:1914-1918(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-98430682; PubMed-9759882;				
RA	Wang R.-F., Johnston S.L., Zeng G., Topalian S.L.,				
RA	Schwarzenberger D.J., Rosenberg S.A.,				
RT	"A breast and melanoma-shared tumor antigen: T cell responses to				
RT	antigenic peptides translated from different open reading frames";				
RL	J. Immunol. 161:3596-3606(1998).				
CC	- TISSUE SPECIFICITY: EXPRESSED IN TESTIS AND OVARY AND IN A WIDE				
CC	VARIETY OF CANCERS. DETECTED IN UTERINE MYOMETRIUM.				
CC	- SIMILARITY: STRONG. TO LAGE-1.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; U87459; AAB49693.1; -				
DR	EMBL; AJ003149; CAA05908.1; -				
DR	EMBL; AF038567; AAD05202.1; -				
DR	MIM; 300156; -				
KW	Transmembrane; Antigen.				
FT	DOMAIN 5 82				
FT	TRANSMEM 156 172				
					GLY-RICH.
					POTENTIAL.

34	139	14.5	747	1	CA12_BOVIN	P02459	bov taurin
35	139	14.5	1027	1	CAF2_RIFPA	P07074	filif pach
36	138	14.4	603	1	YD25_MYCTU	Q10637	mycobacteri
37	138	14.4	1362	1	CA21_CHICK	P02467	gallus gall
38	138	14.4	1464	1	CA13_MOUSE	P08121	mus musculu
39	137.5	14.3	103	1	YD03_HVSC	P22577	hepervirus
40	137.5	14.3	801	1	Y747_MYCTU	Q53810	mycobacteri
41	137.5	14.3	1804	1	CA1B_MOUSE	Q61245	mus musculu
42	137.5	14.3	3067	1	CA1C_MOUSE	Q60847	mus musculu
43	137	14.3	3176	1	CA36_HUMAN	P12111	homo sapien
44	136.5	14.2	1364	1	CA21_BOVIN	P02465	bov taurin
45	136.5	14.2	1366	1	CA21_CANFA	O46392	canis fam11



SO SEQUENCE 180 AA: 17992 MW: B122C5C28B81569 CRC64:

Query Match 100.0%; Score 959; DB 1; Length 180;

Best Local Similarity 100.0%; Pred. No. 5.4e-61;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MOAERGTGSGTGADGPGGPIPDGPGNAGGPGGAGATGGRGPRGAGAAASGPGGGA 60  
 DB 1 MOAERGTGSGTGADGPGGPIPDGPGNAGGPGGAGATGGRGPRGAGAAASGPGGGA 60  
 OY 61 PRGPRGGAASGLNCCGCGARGPESRLEFFYLAMPATMEALARRSLAODAPPLPVP 120  
 DB 61 PRGPRGGAASGLNCCGCGARGPESRLEFFYLAMPATMEALARRSLAODAPPLPVP 120  
 OY 121 VLKFEFTVSGNLTIRLTADHRLQLSISCTLOQLSLMTITQCFPLVFLAOPSGQR 180  
 DB 121 VLKFEFTVSGNLTIRLTADHRLQLSISCTLOQLSLMTITQCFPLVFLAOPSGQR 180

## RESULT 2

LAG1\_HUMAN STANDARD; PRT; 210 AA.

AC 075637; 075637; 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE LAGE-1 protein.

GN CTAG2 OR LAGE1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Melanoma;

RA MEDLINE=98289662; PubMed=9626360;

RA Lethé B., Lucas S., Michaux L., de Smet C., Godelaine D., Serrano A.,

de Placen E., Boon T.;

RT "LAGE-1, a new gene with tumor specificity.";

RL Int. J. Cancer 76:903-908(1998).

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: LAGE-1A AND LAGE-1B (SHOWN

HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: TESTIS AND VERY LOW LEVEL IN PLACENTA AND IN

SOME UTERUS SAMPLES. OBSERVED IN 25-50% OF TUMOR SAMPLES OF

MELANOMAS, NON-SMALL-CELL LUNG CARCINOMAS, BLADDER, PROSTATE AND

HEAD AND NECK CANCERS.

CC -1- DOMAIN: A TRANSMEMBRANE DOMAIN IS PRESENT IN ISOFORM LAGE-1A.

CC -1- SIMILARITY: STRONG, TO NT-ESO-1/CTAG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

FT VARIANT 138 138 W -> R.  
 SO SEQUENCE 210 AA: 21119 MW: 880E00A55588B CRC64;

Query Match 62.0%; Score 595; DB 1; Length 210;

Best Local Similarity 82.0%; Pred. No. 1.9e-35;

Matches 109; Conservative 11; Mismatches 13; Indels 0; Gaps 0;

OY 1 MOAERGTGSGTGADGPGGPIPDGPGNAGGPGGAGATGGRGPRGAGAAASGPGGGA 60  
 DB 1 MOAERGTGSGTGADGPGGPIPDGPGNAGGPGGAGATGGRGPRGAGAAASGPGGGA 60  
 OY 61 PRGPRGGAASGLNCCGCGARGPESRLEFFYLAMPATMEALARRSLAODAPPLPVP 120  
 DB 61 PRGPRGGAASGLNCCGCGARGPESRLEFFYLAMPATMEALARRSLAODAPPLPVP 120  
 OY 121 VLKFEFTVSGNLT 133  
 DB 121 VLKFEFTVSGNLT 133

## RESULT 3

HYDL\_STRCO STANDARD; PRT; 627 AA.

AC P42534; 095217; 01-NOV-1995 (Rel. 32, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative polyketide hydroxylase (EC 1.14.13.-) (WHE ORF VIII).

GN SC659.12C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID:1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D., James K.D., Parkhill J., Barrell B.G.,

RA Rastandean M.A.;

RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RL [2]

RP SEQUENCE OF 1-255 FROM N.A.

RC STRAIN=A3(2);

RA MEDLINE=94075247; PubMed=8253693;

RA Blanco G., Pereda A., Brian P., Mendez C., Chater K.F., Salas J.A.;

RT "A hydroxylase-like gene product contributes to synthesis of a

polyketide spore pigment in Streptomyces halstedii.";

CC -1- FUNCTION: INVOLVED IN DEVELOPMENTALLY REGULATED SYNTHESIS OF A

COMPOUND BIOSYNTHETICALLY RELATED TO POLYKETIDE ANTIBIOTICS,

WHICH IS ESSENTIAL FOR SPORE COLOR IN STREPTOCOCCUS COELICOLOR.

CC -1- SIMILARITY: BELONGS TO THE PHEA/TRDB FAMILY OF FAD MONOOXYGENASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

FT VARIANT 138 138 W -> R.  
 SO SEQUENCE 210 AA: 21119 MW: 880E00A55588B CRC64;

Query Match 62.0%; Score 595; DB 1; Length 210;

Best Local Similarity 82.0%; Pred. No. 1.9e-35;

Matches 109; Conservative 11; Mismatches 13; Indels 0; Gaps 0;

OY 1 MOAERGTGSGTGADGPGGPIPDGPGNAGGPGGAGATGGRGPRGAGAAASGPGGGA 60  
 DB 1 MOAERGTGSGTGADGPGGPIPDGPGNAGGPGGAGATGGRGPRGAGAAASGPGGGA 60  
 OY 61 PRGPRGGAASGLNCCGCGARGPESRLEFFYLAMPATMEALARRSLAODAPPLPVP 120  
 DB 61 PRGPRGGAASGLNCCGCGARGPESRLEFFYLAMPATMEALARRSLAODAPPLPVP 120  
 OY 121 VLKFEFTVSGNLT 133  
 DB 121 VLKFEFTVSGNLT 133

## RESULT 3

HYDL\_STRCO STANDARD; PRT; 627 AA.

AC P42534; 095217; 01-NOV-1995 (Rel. 32, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative polyketide hydroxylase (EC 1.14.13.-) (WHE ORF VIII).

GN SC659.12C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID:1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D., James K.D., Parkhill J., Barrell B.G.,

RA Rastandean M.A.;

RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RL [2]

RP SEQUENCE OF 1-255 FROM N.A.

RC STRAIN=A3(2);

RA MEDLINE=94075247; PubMed=8253693;

RA Blanco G., Pereda A., Brian P., Mendez C., Chater K.F., Salas J.A.;

RT "A hydroxylase-like gene product contributes to synthesis of a

polyketide spore pigment in Streptomyces halstedii.";

CC -1- FUNCTION: INVOLVED IN DEVELOPMENTALLY REGULATED SYNTHESIS OF A

COMPOUND BIOSYNTHETICALLY RELATED TO POLYKETIDE ANTIBIOTICS,

WHICH IS ESSENTIAL FOR SPORE COLOR IN STREPTOCOCCUS COELICOLOR.

CC -1- SIMILARITY: BELONGS TO THE PHEA/TRDB FAMILY OF FAD MONOOXYGENASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 24, 2002, 11:17:18 ; Search time 197.76 Seconds

(without alignments)  
91.347 Million cell updates/sec

Title: PCT-US02-13994-28

Perfect score: 1012

Sequence: 1 MNCDDAFARRPTVGAQIPERK.....RRKQLYVEISDPEDDE 188

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1012	100.0	188	2 S55058	synovial sarcoma X
2	772	76.3	188	2 S55057	SSX1 protein - hum
3	107	10.6	1111	2 T00324	hypothetical prote
4	101	10.0	289	2 A48913	KRAA zinc finger p
5	99	9.8	392	2 T19327	hypothetical prote
6	94.5	9.3	952	2 S64473	translaton initia
7	91.5	9.0	252	2 T20321	hypothetical prote
8	91.5	9.0	1507	2 B47328	natural killer cel
9	90.5	8.9	1153	2 T21386	hypothetical prote
10	88.5	8.7	213	2 U00075	neutromodulin - g
11	88.5	8.7	270	2 B88734	protein F32E10.2
12	88	8.7	445	2 T23908	hypothetical prote
13	87.5	8.6	186	2 S30221	nonhistone chromos
14	87.5	8.6	209	1 NSH082	nonhistone chromos
15	87.5	8.6	541	2 T48836	hypothetical prote
16	87.5	8.6	586	2 B84434	hypothetical prote
17	87	8.6	346	2 C88961	protein P59A7.5
18	87	8.6	1559	2 T30535	calcium channel al
19	86.5	8.5	654	2 A5785	finger protein ZNF
20	86	8.5	1087	1 QPM5H	neurofilament trip
21	85.5	8.4	990	2 I51618	nucleolar phosphop
22	85.5	8.4	2251	2 B54972	voltage-dependent
23	85.5	8.4	2270	2 A54972	nonhistone chromos
24	85	8.4	210	2 A34719	hypothetical prote
25	85	8.4	210	2 S34774	high mobility grou
26	85	8.4	1162	2 T51040	hypothetical prote
27	85	8.4	1167	2 A35066	streptococcal C5a
28	84.5	8.3	755	2 S32103	filensin - bovine
29	84	8.3	1356	2 T16754	hypothetical prote

30	84	8.3	1647	2 S45252	8NF2beta protein -
31	83.5	8.3	301	2 E29149	proline-rich prote
32	83.5	8.3	636	2 I48689	gene NK10 protein
33	83.5	8.3	920	2 T52426	dynamitin-like prote
34	83.5	8.3	1280	2 T00365	hypothetical prote
35	83.5	8.3	1626	2 A39242	DNA topoisomerase
36	83.5	8.3	1663	2 T42092	B-afadin - rat
37	83.5	8.3	1829	2 T41751	1-afadin - rat
38	83	8.2	783	2 F84514	hypothetical prote
39	83	8.2	1062	2 T14151	inv protein - mous
40	83	8.2	1110	2 T51116	NF-180 - sea lamp
41	83	8.2	1235	2 T17457	SAHA protein - Afr
42	83	8.2	1948	2 S00485	gene 11-1 protein
43	82.5	8.2	1069	2 S27922	nuclear antigen EB
44	82	8.1	279	2 T21068	hypothetical prote
45	82	8.1	279	2 T15662	hypothetical prote

#### ALIGNMENTS

RESULT 1  
S55058  
synovial sarcoma X chromosome breakpoint protein SSX2 - human  
C/Species: Homo sapiens (man)  
C/Date: 28-Oct-1996 #sequence-revision 13-Mar-1997 #text-change 01-Dec-2000  
C/Accession: S55058; I68675; S46270  
R/Crew: A.J.: Clark, J.: Fisher, C.: G111, S.: Grimer, R.: Chand, A.: Shipley, J.: Gu  
EMBO J. 14, 2333-2340, 1995  
A/Title: Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homology t  
A/Reference number: S55057; MUID:95292974  
A/Accession: S55058  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-188 <CRE>  
A:Cross-references: EMBL:X86175; NID:9829113; PID:CAA6011.1; PID:q4218445  
R/de Leuw, B.; Bailemans, M.; Olde Weghuis, D.; Geurts van Kessel, A.  
Hum. Mol. Genet. 4, 1097-1099, 1995  
A/Title: Identification of two alternative fusion genes, SYT-SSX1 and SYT-SSX2, in t  
A/Reference number: I54381; MUID:95384157  
A/Accession: I68675  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 111-188 <RES>  
A:Cross-references: GB:S79332; NID:91087049; PID:NAB35379.1; PID:q1087050  
R/Clark, J.: Rocques, P.J.: Crew, A.J.: G111, S.: Shipley, J.: Chan, A.M.L.: Guetero  
Nature Genet. 7, 502-508, 1994  
A/Title: Identification of novel genes, SYT and SSX, involved in the t(X;18) (p11.2;q  
A/Reference number: S46269; MUID:95038836  
A/Accession: S46270  
A/Molecule type: mRNA  
A/Residues: 111-188 <CLA>  
A:Cross-references: EMBL:X79200  
C/Genetics:  
A:Gene: SYT-SSX2

Query Match 100.0%; Score 1012; DB 2; Length 188;  
Best local similarity 100.0%; Pred. No. 1.8e-78;  
Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNCDDAFARRPTVGAQIPERKAFDDIAKFFSEKEMKKASKEIFYYMKRYEAMTK	60
DB	1	MNCDDAFARRPTVGAQIPERKAFDDIAKFFSEKEMKKASKEIFYYMKRYEAMTK	60
QY	61	LGKATLPPFMCNKRAEDPDGNDLNDNRNOVERPOMTFGRLOGISPTIMKRAFEK	120
DB	61	LGKATLPPFMCNKRAEDPDGNDLNDNRNOVERPOMTFGRLOGISPTIMKRAFEK	120
QY	121	NDSEEPVPAASGPDNGKELCPGKPTTSEKIHRSGRGSHAMTHRLRRKQLYVEEI	180
DB	121	NDSEEPVPAASGPDNGKELCPGKPTTSEKIHRSGRGSHAMTHRLRRKQLYVEEI	180

OY 181 SDEEDDE 188  
 |||||  
 Db 181 SDEEDDE 188

## RESULT 2

SSX1 protein - human  
 C:Species: Homo sapiens (man)  
 C>Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Jul-2000  
 C:Accession: S55057; 154381  
 R:Creu, A.J.; Clark, J.; Fisher, C.; Gill, S.; Grimer, R.; Chand, A.; Shipley, J.; Guste  
 EMBL J. 14, 2333-2340, 1995  
 A>Title: Fusion of SYX to two genes, SSX1 and SSX2, encoding proteins with homology to t  
 A:Reference number: S55057; MUID:95292974  
 A:Accession: S55057  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-188 <CR>  
 A:Cross-References: EMBL:X86174; NID:g829112; PIDN:CAA60110.1; PID:g4218444  
 R:de Leeuw, B.; Bailemans, M.; Olde Weghuis, D.; Geurts van Kessel, A.  
 Hum. Mol. Genet. 4, 1097-1099, 1995  
 A>Title: Identification of two alternative fusion genes, SYX-SSX1 and SYX-SSX2, in t(X;1  
 A:Reference number: 154381; MUID:95384157  
 A:Accession: 154381  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 111-188 <RES>  
 A:Cross-References: GB:S79325; NID:g1087047; PIDN:AAB35378.1; PID:g1087048  
 C:Genetics:  
 A:Gene: SYX-SSX1

Query Match 76.38; Score 772; DB 2; Length 188;  
 Best Local Similarity 78.24; Pred. No. 3.5e-58;

Matches 147; Conservative 12; Mismatches 29; Indels 0; Gaps 0;

OY 1 MNGDDAFARPTVGAQIFEXIOKAFDIAIFYSESEKEMKASEKIFYYVMKRYEAMTK 60  
 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11  
 Db 1 MNGDDAFARPTVGAQIFEXIOKAFDIAIFYSESEKEMKASEKIFYYVMKRYEAMTK 60  
 OY 61 LGRFATLTPFCNKRRAEDFQGNLDNDPNRQNOYERDPTFGRLGISTPKIMPKRAEKG 120  
 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11  
 Db 61 LGRFATLTPFCNKRRAEDFQGNLDNDPNRQNOYERDPTFGRLGISTPKIMPKRAEKG 120  
 OY 121 NSEEEVPEASGPNQNGKELCPGKPTTSEKTHERSGPKRGEMHATRLERKQVYIERI 180  
 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11  
 Db 121 NSEEEVPEASGPNQNGKELCPGKPTTSEKTHERSGPKRGEMHATRLERKQVYIERI 180  
 OY 181 SDEEDDE 188  
 |||||  
 Db 181 SDEEDDE 188

## RESULT 3

hypothetical protein KIAA0543 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00324  
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
 DNA Res. 5, 31-39, 1998  
 A>Title: Prediction of the coding sequences of unidentified human genes. IX. The complet  
 A:Reference number: Z14086; MUID:98290545  
 A:Accession: T00324  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1111 <NAG>  
 A:Cross-References: EMBL:AB011115; NID:g3043609; PIDN:BAA25469.1; PID:g3043610  
 C:Genetics:  
 A:Note: KIAA0543

Query Match 10.64; Score 107; DB 2; Length 1111;

Best Local Similarity 24.24; Pred. No. 0.49;

Matches 44; Conservative 34; Mismatches 70; Indels 34; Gaps 9;

OY 16 QIP---EKIQAFDIDIAIFYSESEKEMKASEKIFYYVMKRYEAMRLGKATLPPFM 71  
 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11  
 Db 265 EYVVFEEELPYFEDVAVYFTRFEMGLDROKELRDVNRKNYELLISPAAKAPOLI 324  
 OY 72 --CNRAEDFQGNLDNDPNRQNOYERDPTFGRLGISTPKIMPKRAEKGNDSEVPEA 129  
 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11  
 Db 325 SKLERRAAPW-----IDPN-----GPKNGKGRPG-NKKVAVREADTQASADSALL 372  
 OY 130 SQPQNGKELCPGKPTTSEKTHERSGPKRGEMHATRLERKQVYIERI 183  
 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11  
 Db 373 PGSPVRAASC-----CSSICEGCGDPRIKRTIPRSTIONSWFQGFPLVT-----DP 422  
 OY 184 EE 185  
 ||  
 Db 423 KE 424

## RESULT 4

KRAB zinc finger protein 75 - human  
 N:Alternate names: ZNF75 protein  
 C:Species: Homo sapiens (man)  
 C>Date: 10-May-1996 #sequence\_revision 10-May-1996 #text\_change 05-Nov-1999  
 C:Accession: A48913; A43288; S47339  
 R:Villa, A.; Zucchi, I.; Pilla, G.; Strina, D.; Susani, L.; Morali, F.; Patrosso, C.;  
 Genomics 18, 223-229, 1993  
 A>Title: ZNF75: Isolation of a cDNA clone of the KRAB zinc finger gene subfamily mapp  
 A:Reference number: A48913; MUID:94116987  
 A:Accession: A48913  
 A:Status: preliminary; nucleic acid sequence not shown; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-289 <RES>  
 A:Cross-References: GB:S47370; NID:g460902; PIDN:AAB29696.1; PID:g460903  
 A:Experimental source: Lung fibroblasts  
 R:Villa, A.; Patrosso, C.; Binno, I.; Fratini, A.; Repetto, M.; Mostardini, M.; Eva  
 Genomics 13, 1231-1236, 1992  
 A>Title: Isolation of a zinc finger motif (ZNF75) mapping on chromosome xq26.  
 A:Reference number: A43288; MUID:92372018  
 A:Accession: A43288  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 71-289 <VIL>  
 A:Cross-References: GB:S43109; NID:g254027; PIDN:AAB22971.1; PID:g254028  
 A:Note: sequence extracted from NCBI backbone (NCBIN:111827, NCBI:P111830)  
 R:Marino, M.; Archidiacono, N.; Franze, N.; Rosati, M.; Rocchi, M.; Ballabio, A.; Gri  
 submitted to the EMBL Data Library, July 1992  
 A:Description: The human zinc finger gene family: Isolation, mapping and expression o  
 A:Reference number: S47339  
 A:Accession: S47339  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 139-289 <KAB>  
 C:Genetics:  
 A:Gene: GDB:ZNF75  
 A:Cross-References: GDB:131726; OMIM:314997  
 A:Map position: Xq26-Xq26  
 A:Introns: 11/3  
 C:Keywords: zinc finger

Query Match 10.04; Score 101; DB 2; Length 289;

Best Local Similarity 22.24; Pred. No. 0.33;

Matches 46; Conservative 22; Mismatches 65; Indels 74; Gaps 9;

OY 17 IPEKID-KAFDIDIAIFYSESEKEMKASEKIF-YTMKRYEAMRLGKATLPPFMCK 74  
 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11 11:11 ||||| 11:11  
 Db 7 LPESLSLTFEDVAVYFSEEMGLNLEKTLNDVMDIYIVISGLK----- 58

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 11:24:44 ; Search time 95.25 seconds

(without alignments)  
127.642 Million cell updates/sec

Title: PCT-US02-13994-29

Sequence: 1 MPELRSOMCKPECELEANG.....GGPHISYPLHEWVLRGEE 314

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1636	100.0	314	1	MAG3_HUMAN
2	1580	96.6	314	1	MAG6_HUMAN
3	1401	85.6	314	1	MAGC_HUMAN
4	1367	83.6	314	1	MAG2_HUMAN
5	1076.5	65.8	317	1	MAG4_HUMAN
6	1025.5	63.7	309	1	MAG1_HUMAN
7	920.5	56.3	319	1	MAG8_HUMAN
8	919.5	56.2	315	1	MAG9_HUMAN
9	783.5	47.9	369	1	MAG4_HUMAN
10	695.5	42.5	234	1	MAG8_HUMAN
11	568	34.7	373	1	MAG1_HUMAN
12	556.5	34.0	345	1	MGB4_HUMAN
13	539.5	33.0	347	1	MGB1_HUMAN
14	519	31.7	319	1	MGB2_HUMAN
15	503	30.7	346	1	MGB3_HUMAN
16	485	29.6	1142	1	MGB1_HUMAN
17	447	27.3	124	1	MAG5_HUMAN
18	364	22.2	606	1	MGB2_HUMAN
19	363.5	22.2	775	1	MGB1_HUMAN
20	362.5	22.2	775	1	MGB1_HUMAN
21	356	21.8	529	1	MGB1_HUMAN
22	349	21.3	308	1	MGB1_HUMAN
23	347.5	21.2	778	1	MGB1_HUMAN
24	341	20.8	1387	1	TROP_HUMAN
25	316	19.3	490	1	MGB2_HUMAN
26	281	17.2	321	1	MGB2_HUMAN
27	277	16.9	325	1	MGB2_HUMAN
28	243	14.9	219	1	MGB1_HUMAN
29	103	6.3	1805	1	MGB1_HUMAN
30	96.5	5.9	1202	1	MGB1_HUMAN
31	94.5	5.8	465	1	MGB1_HUMAN
32	94	5.7	563	1	MGB1_HUMAN
33	93.5	5.7	1204	1	MGB1_HUMAN

34	93	5.7	1065	1	SEB4_YEAST
35	92	5.6	575	1	APB3_HUMAN
36	92	5.6	1024	1	POP1_HUMAN
37	91.5	5.6	677	1	CHLD_SYNP7
38	91	5.6	341	1	TEB2_HUMAN
39	91	5.6	380	1	GBA1_ORF5A
40	89.5	5.5	3519	1	OL56_STRAT
41	89	5.4	1411	1	TCOF_HUMAN
42	89	5.4	1607	1	MIPR_LYMST
43	88.5	5.4	528	1	TY3H_HUMAN
44	88.5	5.4	1204	1	MGB1_HUMAN
45	88.5	5.4	1527	1	CAH_MOUSE

#### ALIGNMENTS

RESULT	ID	STANDARD	PRT	AA
1	MAG3_HUMAN			
AC	P43357			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Melanoma-associated antigen 3 (MAGE-3 antigen) (Antigen M22-D).			
GN	MAGE3 OR MAGE3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND MUTAGENESIS.			
RC	TISSUE=Blood;			
RC	MEDLINE-94157413; PubMed-8113684;			
RA	Gautier J., van den Eynde B., van der Brugge P., Romero P.,			
RA	Gautier J., de Plaeen E., Leche B., Brasseur F., Boon T.;			
RT	"Human gene MAGE-3 codes for an antigen recognized on a melanoma by			
RT	autologous cytolytic T lymphocytes.";			
RT	J. Exp. Med. 179:921-930(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skin;			
RC	MEDLINE-94311935; PubMed-8037761;			
RA	Ding M., Beck R.J., Keller C.J., Fenton R.G.;			
RT	"Cloning and analysis of MAGE-1-related genes.";			
RT	Biochem. Biophys. Res. Commun. 202:549-555(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE-20314869; PubMed-10854409;			
RA	Mallon A.M., Platzner M., Bates R., Gloeckner G., Botcherby M.,			
RA	Nordstiek G., Strivens M.A., Kioschis P., Dangel A., Cunningham D.,			
RA	Straw R., Weston P., Hunter C., Gilbert M., Fernando S., Goodall K.,			
RA	Kerry G., Greystrong J.S., Clark D., Goerdes M., Blechschmidt K.,			
RA	Rump A., Hinzmann B., Mundy C.R., Miller W., Pousta A., Helman G.E.,			
RA	Rhodes M., Denny P., Rosenthal A., Brown S.D.M.;			
RT	"Comparative genome sequence analysis of the Bpa/Str region in mouse			
RT	and man.";			
RT	Genome Res. 10:758-775(2000).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lung;			
RC	Strausberg R.;			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: NOT KNOWN, THOUGH MAY PLAY A ROLE IN EMBRYONAL			
CC	DEVELOPMENT AND TUMOR TRANSFORMATION OR ASPECTS OF TUMOR			
CC	PROGRESSION. ANTIGEN RECOGNIZED ON A MELANOMA BY AUTOLOGOUS			
CC	CYTOLYTIC T LYMPHOCYTES.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MANY TUMORS OF SEVERAL TYPES,			
CC	SUCH AS MELANOMA, HEAD AND NECK SQUAMOUS CELL CARCINOMA, LUNG			
CC	CARCINOMA AND BREAST CARCINOMA, BUT NOT IN NORMAL TISSUES EXCEPT			
CC	FOR TESTES AND PLACENTA. NEVER EXPRESSED IN KIDNEY TUMORS,			
CC	LEUKEMIAS AND LYMPHOMAS.			
CC	-1- SIMILARITY: CONTAINS 1 MAGE DOMAIN.			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----  
 DR EMBL: U03735; AAA17446.1;  
 DR EMBL: U82671; AAF44791.1;  
 DR EMBL: BC000340; AAH00340.1;  
 DR MIM: 300174;  
 DR InterPro: IPR002190; MAGE.  
 DR Pfam: PF01454; MAGE.1.  
 DR PROSITE: PS50838; MAGE.1.  
 KW Antigen; Multigene family; Tumor antigen.  
 FT DOMAIN 109 308 MAGE.  
 FT DOMAIN 40 43 POLY-SER.  
 FT MUTAGEN 170 170 D->A: ABOLISHES HLA-A1 BINDING.  
 FT MUTAGEN 176 176 Y->A: ABOLISHES HLA-A1 BINDING.  
 FT SEQUENCE 314 AA: 34747 MW: 355813 DIC946A1 CRC64.

Query Match 100.0%; Score 1636; DB 1; Length 314;  
 Best Local Similarity 100.0%; Pred. No. 1,3e-126;  
 Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MPLEORSCHCKPEGLERARGLVGAQAPATEBOEAASSSTLVETVLGEVPAAESPD 60  
 DB 1 MPLEORSCHCKPEGLERARGLVGAQAPATEBOEAASSSTLVETVLGEVPAAESPD 60  
 OY 61 PROSGASSLPTTNNYPLMSOSYEDSSNOEEBSPSTFPDLESEFOALS SRKVAELVHFL 120  
 DB 61 PROSGASSLPTTNNYPLMSOSYEDSSNOEEBSPSTFPDLESEFOALS SRKVAELVHFL 120  
 OY 121 LKTYRAREPTVTAEMLGSVGNMOTFFPVIFSKASSSLQVFGIEIEMVDPIGHYIYFAT 180  
 DB 121 LKTYRAREPTVTAEMLGSVGNMOTFFPVIFSKASSSLQVFGIEIEMVDPIGHYIYFAT 180  
 OY 181 CIGLSYDGLGDNQIMPKAGLIIYLAIIAREGDCAPEEKIMEELSVLEFEGREDSTILG 240  
 DB 181 CIGLSYDGLGDNQIMPKAGLIIYLAIIAREGDCAPEEKIMEELSVLEFEGREDSTILG 240  
 OY 241 DPKLLTQHFVQENTLEYRQVPGSDPACYEFLMGPRALVETSYVKVLHVMKISGSPHIS 300  
 DB 241 DPKLLTQHFVQENTLEYRQVPGSDPACYEFLMGPRALVETSYVKVLHVMKISGSPHIS 300  
 OY 301 YPPLHEWVLRGEE 314  
 DB 301 YPPLHEWVLRGEE 314

RESULT 2  
 MAG6\_HUMAN STANDARD; PRT; 314 AA.  
 ID MAG6\_HUMAN  
 AC PA3360;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Melanoma-associated antigen 6 (MAGE-6 antigen) (MAGE3B).  
 GN MAGE6 OR MAGE6.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95012457; PubMed=7927340;  
 RA de Plien E., Arden K., Traversari C., Gaforio J.J., Szikora J.-P.,  
 RA Brasseur R., Brasseur F., van der Bruggen P., Ieche B., Lurquin C.,  
 RA de Smet C., Chomez P., de Backer O., Cavenee W., Boon T.;  
 RA Structure, chromosomal localization, and expression of 12 genes of

RT the MAGE family.\*;  
 RL Immunogenetics 40:360-369(1994).

RN [2]  
 RP SEQUENCE FROM N.A.

RC TISSUE-Skin;  
 RX MEDLINE=94311935; PubMed=8037761;

RA Ding M., Beck R.J., Keller C.J., Fenton R.G.;  
 RT Cloning and analysis of MAGE-1-related genes.\*;  
 RL Blochem. Biophys. Res. Commun. 202:549-555(1994).

RN [3]  
 RP SEQUENCE FROM N.A.

RA MEDLINE=95369706; PubMed=7642112;  
 RA Imai Y., Shichijo S., Yamada A., Katayama T., Yano H., Itoh K.;

RT Sequence analysis of the MAGE gene family encoding human tumor-  
 rejection antigens.\*;  
 RL Gene 160:287-290(1995).

CC -1- FUNCTION: NOT KNOWN, THOUGH MAY PLAY A ROLE IN TUMOR  
 OR ASPECTS OF TUMOR PROGRESSION.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY TUMORS OF SEVERAL TYPES,  
 SUCH AS MELANOMA, HEAD AND NECK SQUAMOUS CELL CARCINOMA, LUNG

CC CARCINOMA AND BREAST CARCINOMA, BUT NOT IN NORMAL TISSUES EXCEPT  
 FOR TESTES.

CC -1- SIMILARITY: CONTAINS 1 MAGE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----

DR EMBL: U10691; AA68875.1;  
 DR EMBL: U10339; AA19006.1;  
 DR EMBL: D32076; BAA06842.1;  
 DR MIM: 300176;  
 DR InterPro: IPR002190; MAGE.  
 DR Pfam: PF01454; MAGE.1.  
 DR PROSITE: PS50838; MAGE.1.  
 KW Antigen; Multigene family; Tumor antigen.  
 FT DOMAIN 109 308 MAGE.  
 FT DOMAIN 40 43 POLY-SER.  
 FT SEQUENCE 314 AA: 34891 MW: 298833 C/FAGE50263 CRC64.

Query Match 96.6%; Score 1580; DB 1; Length 314;  
 Best Local Similarity 95.9%; Pred. No. 4.9e-122;  
 Matches 301; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 MPLEORSCHCKPEGLERARGLVGAQAPATEBOEAASSSTLVETVLGEVPAAESPD 60  
 DB 1 MPLEORSCHCKPEGLERARGLVGAQAPATEBOEAASSSTLVETVLGEVPAAESPD 60  
 OY 61 PROSGASSLPTTNNYPLMSOSYEDSSNOEEBSPSTFPDLESEFOALS SRKVAELVHFL 120  
 DB 61 PROSGASSLPTTNNYPLMSOSYEDSSNOEEBSPSTFPDLESEFOALS SRKVAELVHFL 120  
 OY 121 LKTYRAREPTVTAEMLGSVGNMOTFFPVIFSKASSSLQVFGIEIEMVDPIGHYIYFAT 180  
 DB 121 LKTYRAREPTVTAEMLGSVGNMOTFFPVIFSKASSSLQVFGIEIEMVDPIGHYIYFAT 180  
 OY 181 CIGLSYDGLGDNQIMPKAGLIIYLAIIAREGDCAPEEKIMEELSVLEFEGREDSTILG 240  
 DB 181 CIGLSYDGLGDNQIMPKAGLIIYLAIIAREGDCAPEEKIMEELSVLEFEGREDSTILG 240  
 OY 241 DPKLLTQHFVQENTLEYRQVPGSDPACYEFLMGPRALVETSYVKVLHVMKISGSPHIS 300  
 DB 241 DPKLLTQHFVQENTLEYRQVPGSDPACYEFLMGPRALVETSYVKVLHVMKISGSPHIS 300  
 OY 301 YPPLHEWVLRGEE 314  
 DB 301 YPPLHEWVLRGEE 314